
Beyond Euclidean Summaries: Online Change Point Detection for Distribution-Valued Data

Anonymous Authors¹

Abstract

Existing online change-point detection (CPD) methods rely on fixed-dimensional Euclidean summaries, implicitly assuming that distributional changes are well captured by moment-based or feature-based representations. They can obscure important changes in distributional shape or geometry. We propose a geometry-aware CPD framework that treats streaming batch data as a stochastic process on the 2-Wasserstein space. Our method detects changes in the law of this process by mapping each empirical distribution to a tangent space relative to a pre-change Fréchet barycenter, yielding a reference-centered local linearization of 2-Wasserstein space. This representation enables sequential detectors by adapting classical multivariate monitoring statistics to tangent fields. We provide theoretical guarantees and demonstrate, via synthetic and real-world experiments, that our approach detects complex distributional shifts with reduced detection delay at matched ARL_0 compared with moments-based and model-free baselines.

1. Introduction

Online change point detection (CPD) aims to detect an abrupt change in a sequential data stream as quickly as possible while controlling false alarms. Formally, one constructs a stopping time τ and seeks small detection delay under post-change regimes subject to a lower bound on the no-change average run length $ARL_0 = \mathbb{E}_\infty[\tau]$ (Page, 1954; Truong et al., 2020). Online CPD is a fundamental tool for monitoring nonstationarity and distribution shift in deployed learning systems, with applications spanning large-scale Machine Learning (ML) services, healthcare and bioinformatics, finance, and social platforms (Aminikhang-

¹Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

hahi & Cook, 2017; Aminikhanghahi et al., 2018; Truong et al., 2019).

Most online CPD methodologies are formulated for point-valued streams $\{x_t\}$ in Euclidean space. In model-free settings, a common construction is to compare “past” and “recent” segments via a sequential two-sample statistic (often kernel-based) and to stop when this statistic exceeds a threshold (Harchaoui et al., 2008; Li et al., 2015; Chang et al., 2019). More recent work targets high-dimensional streams through computationally efficient, constant-memory recursions (e.g., random-feature approximations and multi-timescale exponential averages) (Keriven et al., 2020). At the same time, CPD has been extended beyond Euclidean vectors to random objects located on manifolds or more general metric spaces (Wang et al., 2013; Bouchard et al., 2020; Dubey & Müller, 2020; Wang et al., 2024).

However, many modern sensing and observational systems produce batches of observations at each time step (e.g., thousands of cells in flow cytometry, or a burst of user-generated content). The natural representation of the observation at time t is the empirical measure $\mu_t = \frac{1}{n_t} \sum_{j=1}^{n_t} \delta_{x_{t,j}}$, and CPD becomes testing for changes in the law of a distribution-valued process $\{\mu_t\}_{t \geq 1}$ (Horváth et al., 2021). In this regime, one may either (i) compare batches via sequential two-sample discrepancies, or (ii) monitor a vector-valued functional of μ_t obtained from moments or learned feature maps. These strategies effectively test changes in selected functionals of μ_t . Consequently, they can have limited power when the departure is predominantly geometric, *i.e.*, when probability mass is redistributed through deformations of the support (translations, shape changes, multimodal reweighting) rather than through changes in low-order moments. This motivates treating each batch as a distribution-valued observation and modeling $\{\mu_t\}$ intrinsically on $(\mathcal{P}_2(\mathcal{X}), W_2)$, where the Wasserstein metric directly quantifies such mass-transport (geometric) changes.

We therefore treat $\{\mu_t\}$ as a stochastic process taking values in the metric space $(\mathcal{P}_2(\mathcal{X}), W_2)$. Under quadratic cost, the 2-Wasserstein distance metrizes discrepancies in mass displacement and is sensitive to changes in support and shape that are not well characterized by low-order moments (Panaretos & Zemel, 2019). From a computational

perspective, modern optimal transport (OT) solvers (*e.g.*, entropic regularization and Sinkhorn iterations) make these geometric comparisons tractable at scale (Cuturi, 2013). For sequential monitoring, however, one must further construct a representation that (i) respects the underlying Wasserstein geometry, (ii) lives in a common coordinate system across time, and (iii) admits principled calibration of false-alarm control.

We address this by anchoring at a pre-change Fréchet barycenter $\bar{\mu}$ and mapping each μ_t to a tangent space through OT. Under mild regularity, the Brenier map yields the radial identity $W_2^2(\bar{\mu}, \mu_t) = \|T_{\bar{\mu}}^{\mu_t} - \text{Id}\|_{L^2(\bar{\mu})}^2$, which furnishes a local linearization of \mathcal{W}_2 at $\bar{\mu}$ in the Hilbert space $L^2(\bar{\mu}; \mathbb{R}^d)$. We then build an online detector by applying multivariate functional PCA with two classical quadratic monitoring statistics (Hotelling’s T^2 and SPE), using a pre-change calibration step to set thresholds targeting a desired ARL_0 .

Contributions. We propose an intrinsic distribution-valued CPD (IDD) scheme with the following contributions:

1. We formulate online CPD for a distribution-valued stochastic process $\{\mu_t\}_{t \geq 1}$ and detect changes in its law by transporting each μ_t to a pre-change Fréchet barycenter $\bar{\mu}$, thereby working directly with Wasserstein geometry rather than moment-based or *ad hoc* Euclidean summaries.
2. We develop a two-chart online monitoring procedure by combining Multivariate Functional PCA (MFPCA) with Hotelling’s T^2 and Squared Prediction Error (SPE) in the Wasserstein tangent space, providing a scalable detector with data-driven threshold calibration.
3. We provide theoretical guarantees and demonstrate performance on extensive simulations and two real-world case studies (FlowCAP abnormal-cell detection and public opinion monitoring on Reddit), where distributional geometry shifts are central. Across these settings, the proposed approach consistently achieves strong detection power for distributional and dependence-geometry perturbations while outperforming moment-based or model-free online CPD approaches.

2. Related Work

Statistical Process Control. The foundation of statistical process control (SPC) was established by Shewhart (1931), who introduced control charts for monitoring process stability using summary statistics such as sample mean and range. While simple and interpretable, this framework compresses data into low-dimensional statistics, thereby missing

shifts in the overall distribution shape. Multivariate extensions have been developed to handle vector-valued processes. Crosier (1988) developed multivariate Cumulative Sum (CUSUM) procedures, while Nelson (1984) introduced supplementary pattern tests for Shewhart charts. However, these rules do not generalize naturally to distribution-valued data. Lowry & Montgomery (1995) provided a review of multivariate SPC methods, noting their limitations in handling non-Euclidean or high-dimensional data. Crucially, Woodall & Montgomery (1999) identified the challenges caused by the shift of distributions in pre-change phase. Our work addresses this gap by directly monitoring distributions in the Wasserstein space, capturing geometric shifts that summary-based charts would miss.

Statistics for Distribution-valued Samples. Optimal transport provides the geometric foundation for our work. Panaretos & Zemel (2019) provided a comprehensive review of Wasserstein statistics. Bhattacharya & Patrangenaru (2005) established the asymptotic theory for Fréchet means on manifolds. Petersen & Müller (2016) proposed transforming densities to a Hilbert space for regression analysis, and Bigot et al. (2017) developed geodesic principal component analysis (PCA) in Wasserstein space, though the latter is computationally demanding. While Chernozhukov et al. (2017) introduced OT-based ranks and Sommerfeld & Munk (2018) derived inference for empirical Wasserstein distances, these works focus largely on static analysis or batch testing. Similarly, energy statistics (Székely & Rizzo, 2013) and kernel two-sample tests (Gretton et al., 2012) are powerful for hypothesis testing but lack the sequential control limits required for CPD. Our contribution is to operationalize these geometric insights into a CPD scheme.

Change-Point Detection. Our work builds upon the expanding literature on change-point detection (CPD) for non-Euclidean data. Xie et al. (2021) provides a comprehensive review of classical sequential detection methods. For metric-valued data that belongs to a metric space, Dubey & Müller (2020) established a framework for Fréchet change-point detection, while Zhang et al. (2025) and Jiang et al. (2024) developed self-normalization based inference for object-valued time series, respectively. Specific to probability distributions, Horváth et al. (2021) proposed a sequential monitoring procedure based on L^2 -Wasserstein distances. However, these approaches often rely on scalar distances or operate in univariate settings, ignoring the direction of distributional shift.

3. Intrinsic Distribution-valued CPD

3.1. Preliminaries and Assumptions

At each time index $t \in \{1, 2, \dots\}$ we observe online streaming batch data, where a batch of N_t samples

$\{x_{t,1}, \dots, x_{t,N_t}\} \subset \mathcal{X} \subset \mathbb{R}^d$. We represent the batch by its empirical measure

$$\mu_t := \frac{1}{N_t} \sum_{j=1}^{N_t} \delta_{x_{t,j}} \in \mathcal{P}_2(\mathcal{X}),$$

and view $\{\mu_t\}_{t \geq 1}$ as a distribution-valued time series sampled from a stochastic process taking values in the metric space $(\mathcal{P}_2(\mathcal{X}), W_2)$. Let $\mathcal{P}(\mathcal{X})$ denote the set of Borel probability measures on \mathcal{X} and define

$$\mathcal{P}_2(\mathcal{X}) := \left\{ \mu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} \|x\|^2 \mu(dx) < \infty \right\}.$$

For $\mu, \nu \in \mathcal{P}_2(\mathcal{X})$, the squared 2-Wasserstein distance in Monge form is

$$W_2^2(\mu, \nu) := \inf_T \left\{ \int_{\mathcal{X}} \|x - T(x)\|^2 \mu(dx) : T : \mathcal{X} \rightarrow \mathcal{X} \text{ Borel, } T_{\#}\mu = \nu \right\}, \quad (1)$$

where $T_{\#}\mu$ denotes the pushforward measure, *i.e.*, $T_{\#}\mu(B) = \mu(T^{-1}(B))$ for all Borel sets $B \subset \mathcal{X}$. When an optimal map exists, Eq. (1) coincides with the *Kantorovich* formulation:

$$W_2^2(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{X}} \|x - y\|^2 \pi(dx, dy), \quad (2)$$

where $\Pi(\mu, \nu)$ denotes the set of couplings with marginals μ and ν .

Assumption 3.1 (Within-batch sampling). For each t , conditional on an underlying distribution $\nu_t \in \mathcal{P}_2(\mathcal{X})$, the observations within the batch are i.i.d.: $x_{t,1}, \dots, x_{t,N_t} \stackrel{iid}{\sim} \nu_t$.

We are interested in detecting the potential change in the distribution-valued sequence $\{\mu_t\}_{t=1}^n$ online, formulated as:

$$H_0 : \mu_t \stackrel{iid}{\sim} \mathbb{P}_1, \quad t = 1, \dots, n, \quad (3)$$

where \mathbb{P}_1 is the common distribution of the samples (*i.e.*, the parent law) on the space of measures $\mathcal{P}_2(\mathcal{X})$. Under the alternative, there exists an unknown change point $\kappa \in \{1, \dots, n\}$ such that

$$t < \kappa, \mu_t \sim \mathbb{P}_1, t \geq \kappa, \mu_t \sim \mathbb{P}_2, \text{ where } \mathbb{P}_1 \neq \mathbb{P}_2. \quad (4)$$

For clarity of exposition, Eq. (4) is stated for a single change point. However, the proposed detection method can naturally detect multiple change points online.

We assume a pre-change period of the length $M \geq 1$ of the process, *i.e.*, $\mu_1, \dots, \mu_M \stackrel{iid}{\sim} \mathbb{P}_1$. Let $\bar{\mu}$ denote the Fréchet barycenter of the measures in the pre-change calibration phase:

$$\bar{\mu} \in \arg \min_{\nu \in \mathcal{P}_2(\mathcal{X})} \frac{1}{M} \sum_{t=1}^M W_2^2(\nu, \mu_t). \quad (5)$$

Assumption 3.2 (Regularity of the reference measure). $\bar{\mu}$ is absolutely continuous with respect to Lebesgue measure.

Remark 3.3. Under quadratic cost $c(x, y) = \|x - y\|^2$, Assumption 3.2 implies that for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ there exists an optimal transport map $T_{\bar{\mu} \rightarrow \mu}$ pushing $\bar{\mu}$ to μ , which is $\bar{\mu}$ -a.e. unique (Brenier's theorem). Moreover, the McCann interpolation

$$\mu_s := ((1-s)\text{Id} + sT_{\bar{\mu} \rightarrow \mu})_{\#}\bar{\mu}, \quad s \in [0, 1],$$

defines the unique W_2 -geodesic from $\bar{\mu}$ to μ (Brenier, 1991; Villani et al., 2008). In our implementation, μ_t is empirical and we compute a Kantorovich plan and its barycentric projection (see Remark 3.5).

3.2. Local Geometric Linearization via Optimal Transport

To enable sequential inference on distribution-valued data, we map the empirical measure μ_t into a common Hilbert space by representing it through the optimal-transport displacement from the pre-change Fréchet barycenter $\bar{\mu}$.

Proposition 3.4 (Radial Isometry, Theorem 8.6 (Villani, 2009)). *Assume $\bar{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$ is absolutely continuous with respect to Lebesgue measure, and consider the quadratic cost $c(x, y) = \|x - y\|^2$. Then for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ there exists a $\bar{\mu}$ -a.e. unique optimal transport map $T_{\bar{\mu}}^\mu$ such that $(T_{\bar{\mu}}^\mu)_{\#}\bar{\mu} = \mu$. Defining the logarithm map in the tangent space*

$$v(\mu) := T_{\bar{\mu}}^{\mu} - \text{Id} \in L^2(\bar{\mu}; \mathbb{R}^d),$$

we have the identity

$$W_2^2(\bar{\mu}, \mu) = \int_{\mathcal{X}} \|v(\mu)(x)\|^2 d\bar{\mu}(x) = \|v(\mu)\|_{L^2(\bar{\mu})}^2. \quad (6)$$

Proposition 3.4 establishes that the Wasserstein distance from the barycenter $\bar{\mu}$ is exactly preserved by the $L^2(\bar{\mu})$ norm of the OT displacement field. Motivated by this radial isometry, we represent each distribution-valued sample μ_i by its linearized embedding, *i.e.*, OT displacement field:

$$v_t := T_{\bar{\mu}}^{\mu_t} - \text{Id} \in L^2(\bar{\mu}; \mathbb{R}^d), \quad (7)$$

Geometrically, v_t corresponds to the Wasserstein logarithm map at reference measure $\bar{\mu}$, projecting the measure μ_t onto the tangent space at $\bar{\mu}$. This transformation maps the manifold $\mathcal{P}_2(\mathcal{X})$ into the tangent space $H := L^2(\bar{\mu}; \mathbb{R}^d)$, allowing us to perform linear statistical analysis in the Hilbert space. Consequently, the problem of monitoring distributional shifts is transformed to monitoring a sequence of multivariate functional data, enabling standard covariance operators, functional analysis, and Hilbert-space statistics to be constructed directly on $\{v_t\}$, avoiding distributional information loss with Euclidean summaries.

Remark 3.5. In practice, observed batches are discrete, where a deterministic Monge map may not exist. We therefore solve the Kantorovich relaxation to obtain an optimal coupling $\pi_t \in \Pi(\bar{\mu}, \mu_t)$ and recover a map via the barycentric projection:

$$T_t^\pi(x) := \mathbb{E}_{\pi_t}[Y | X = x] = \int y \pi_t(dy|x). \quad (8)$$

We then define the tangent field as $v_t(x) := T_t^\pi(x) - x$. This projection provides a canonical vector field in $L^2(\bar{\mu}; \mathbb{R}^d)$ that coincides with the true Monge map whenever the latter exists. For simplicity, we henceforth denote this projected field simply as v_t .

To quantify the approximation gap when using a plan-based projection instead of a Monge map, we provide the following variance decomposition, which is a direct application of the L^2 -projection property of conditional expectation (Doob, 1953; Villani, 2003).

Proposition 3.6 (Barycentric projection variance decomposition). *Let $\mu, \nu \in \mathcal{P}_2(\mathcal{X})$ and let $\pi \in \Pi(\mu, \nu)$ be any coupling with $(X, Y) \sim \pi$. Define the barycentric projection $T^\pi(x) := \mathbb{E}_\pi[Y | X = x]$ and the displacement field $v^\pi(x) := T^\pi(x) - x \in L^2(\mu; \mathbb{R}^d)$. Then the following decomposition holds:*

$$\begin{aligned} \int_{\mathcal{X}} \|x - T^\pi(x)\|^2 \mu(dx) &= \int_{\mathcal{X} \times \mathcal{X}} \|x - y\|^2 \pi(dx, dy) \\ &\quad - \int_{\mathcal{X}} \mathbb{E}_\pi[\|Y - T^\pi(X)\|^2 | X = x] \mu(dx) \end{aligned} \quad (9)$$

and in particular

$$\|v^\pi\|_{L^2(\mu)}^2 = \int_{\mathcal{X}} \|x - T^\pi(x)\|^2 \mu(dx) \quad (10)$$

$$\leq \int_{\mathcal{X} \times \mathcal{X}} \|x - y\|^2 \pi(dx, dy). \quad (11)$$

If π^* is an optimal coupling for $W_2^2(\mu, \nu)$, then $\|v^{\pi^*}\|_{L^2(\mu)}^2 \leq W_2^2(\mu, \nu)$, with equality if and only if π^* is induced by a deterministic map (equivalently, $\text{Var}_{\pi^*}(Y | X) = 0$ μ -a.e.).

3.3. The Optimal Transport-based Detector

Having mapped the distribution-valued stream data to the reference tangent space, we formulate the detection task as monitoring a sequence of multivariate functional observations. In online CPD for high-dimensional and functional streams, the standard paradigm is subspace-based detection. This strategy projects the functional observation onto a low-dimensional principal subspace (e.g., via Multivariate Functional PCA (MFPCA)) (Kuncheva, 2011; Skubalska-Rafajłowicz, 2013; Bakdi & Kouadri, 2017). Motivated by this paradigm, we apply MFPCA to the displacement fields

to obtain a finite-dimensional representation and orthogonal residual component, which produces an OT-based detector built from score-space statistic (functional Hotelling T^2) and a complementary residual-energy statistic (Squared Prediction Error (SPE)) on the tangent space.

Let $\{v_t\}_{t=1}^{n_0} \subset H := L^2(\bar{\mu}; \mathbb{R}^d)$ denote a reference sample of n_0 tangent vector fields collected from the pre-change process. Define the pre-change sample mean $\bar{v} = \frac{1}{n_0} \sum_{t=1}^{n_0} v_t$ and centered fields $\tilde{v}_t = v_t - \bar{v}$. The empirical covariance operator $\hat{C} : H \rightarrow H$ is

$$\begin{aligned} (\hat{C}f)(\cdot) &= \frac{1}{n_0 - 1} \sum_{t=1}^{n_0} \langle f, \tilde{v}_t \rangle_H \tilde{v}_t(\cdot) \\ &= \frac{1}{n_0 - 1} \sum_{t=1}^{n_0} (\tilde{v}_t \otimes \tilde{v}_t) f, \end{aligned} \quad (12)$$

where $(a \otimes b)f := \langle f, b \rangle_H a$.

Due to the finite rank of the empirical covariance operator \hat{C} , we employ a spectral decomposition strategy to regularize the inversion, justified by the Karhunen-Loève theorem. Let $\{(\hat{\lambda}_m, \hat{\phi}_m)\}_{m \geq 1}$ denote the eigenpairs derived from the pre-change reference sample, ordered such that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq 0$, with orthonormal eigenfunctions $\{\hat{\phi}_m\}$ spanning the principal directions of variation in H . For any newly observed tangent vector field v_t from the online stream, we compute its centered displacement $\Delta_t := v_t - \bar{v}$ relative to the pre-change mean. By projecting Δ_t onto the eigenbasis, we obtain the scores $\hat{\xi}_{tm} := \langle \Delta_t, \hat{\phi}_m \rangle_H$. By truncating the Karhunen-Loève expansion to the leading K dimensions, we construct the functional Hotelling's T^2 statistic for the principal subspace:

$$T_{t,K}^2 = \sum_{m=1}^K \frac{\hat{\xi}_{tm}^2}{\hat{\lambda}_m}. \quad (13)$$

Remark 3.7. In the above MFPCA decomposition, the empirical covariance operator \hat{C} maps any linear combination of the centered displacement \tilde{v}_t back to the closure of the gradient subspace in H . Consequently, its eigenfunctions $\{\hat{\phi}_m\}_{m \geq 1}$ also lie in this same subspace. Each $\hat{\phi}_m$ is again a gradient field, so the principal component basis does not leave the tangent space $T_{\bar{\mu}}\mathcal{P}_2(\mathcal{X})$.

Simultaneously, we monitor residuals orthogonal to this subspace using the SPE, also known as the Q -statistic (Jackson & Mudholkar, 1979):

$$\text{SPE}_t = |(I - \hat{P}_K)\Delta_t|_H^2 = \sum_{m > K} \hat{\xi}_{tm}^2, \quad (14)$$

where \hat{P}_K denotes the orthogonal projection onto $\text{span}\{\hat{\phi}_1, \dots, \hat{\phi}_K\}$.

Formally, the functional Hotelling’s T^2 statistic can be written as $T_t^2 = \langle \widehat{C}^\dagger \Delta_t, \Delta_t \rangle_H$, where \widehat{C}^\dagger is the Moore–Penrose pseudoinverse of \widehat{C} . The computable expression Eq. (13) follows from the spectral decomposition of \widehat{C} (see Appendix E).

3.4. Theoretical Properties and Threshold Calibration

To derive the asymptotic behavior of the monitoring statistics, we introduce a regularity condition on the distributional generation process. Motivated by Central Limit Theorems for Fréchet means on manifolds (Bhattacharya & Bhattacharya, 2008; Mattingly et al., 2023), we model the pre-change tangent fields as a Gaussian random field.

Assumption 3.8 (Pre-change Gaussian Structure). The pre-change tangent vector fields $\{v_t\}_{t \leq n_0}$ are independent and identically distributed (i.i.d.) realizations of a mean-zero Gaussian Random Element V in the Hilbert space $H = L^2(\bar{\mu}; \mathbb{R}^d)$. The process is fully characterized by a compact, positive trace-class covariance operator $\Gamma : H \rightarrow H$, with eigenvalues $\lambda_1 > \lambda_2 > \dots > 0$.

Under Assumption 3.8, the functional Hotelling T^2 statistic defined in Eq. (13) follows a known limiting distribution.

Proposition 3.9 (Asymptotic Null Distribution). *Let the pre-change sample size $n_0 \rightarrow \infty$. Under standard consistency conditions for empirical functional eigenpairs (Hall & Hosseini-Nasab, 2006), for any fixed truncation level K , the online Hotelling statistic $T_{t,K}^2$ for a new pre-change observation converges in distribution to a chi-squared random variable:*

$$T_{t,K}^2 \xrightarrow{d} \chi_K^2. \quad (15)$$

Simultaneously, the residual statistic SPE_t converges to a weighted sum of independent chi-squared variables, $\sum_{m>K} \lambda_m Z_m^2$, where $Z_m \sim \mathcal{N}(0, 1)$.

Proof. See Appendix C. \square

Calibration Strategy. While Proposition 3.9 suggests a parametric threshold $h_{T^2} = \chi_{K,1-\alpha}^2$, real-world distribution-valued streams may exhibit non-Gaussian tail behavior. To ensure robustness, we adopt a non-parametric calibration approach. We set the detection thresholds h_{T^2} and h_{SPE} as the empirical $(1 - \alpha)$ quantiles of the pre-change statistics $\{T_{t,K}^2\}_{t=1}^{n_0}$ and $\{\text{SPE}_t\}_{t=1}^{n_0}$. This data-driven strategy automatically accounts for finite-sample effects and the intractable infinite sum of eigenvalues in the SPE limit.

For distribution-valued streams, the null law of $(T_{t,K}^2, \text{SPE}_t)$ is not available in closed form because it depends on the unknown pre-change distribution of tangent fields induced by the OT construction. Once the

stream has been mapped into the monitoring statistics and a fixed threshold pair $(h_{T^2}, h_{\text{SPE}})$ has been chosen, sequential false-alarm control reduces to controlling the one-step exceedance probability under \mathbb{P}_∞ . The following theorem makes this reduction explicit (under i.i.d. monitoring statistics), and the corollary links our empirical-quantile calibration to an explicit finite-sample ARL_0 lower bound.

Theorem 3.10 (Sequential false-alarm control under fixed thresholds). *Let $(h_{T^2}, h_{\text{SPE}})$ be fixed thresholds. Under the assumption that the statistics are i.i.d. during the monitoring phase, the run-length τ follows a geometric distribution with success probability $p_\infty = \mathbb{P}_\infty(T_{1,K}^2 > h_{T^2} \cup \text{SPE}_1 > h_{\text{SPE}})$. Consequently, the global Average Run Length is given by:*

$$\text{ARL}_0 = n_0 + 1 + \frac{1}{p_\infty} \geq n_0 + 1 + \frac{1}{\alpha_{T^2} + \alpha_{\text{SPE}}}. \quad (16)$$

Corollary 3.11 (Empirical-quantile calibration implies ARL control). *If thresholds are set as the $(1 - \alpha)$ empirical quantiles from a calibration sample of size n_0 , the marginal exceedance probabilities for an independent monitoring-phase observation satisfy $\mathbb{P}_\infty(\cdot > h) \leq \alpha + \frac{1}{n_0+1}$. Substituting this into Theorem 3.10 yields the finite-sample guarantee:*

$$\text{ARL}_0 \geq n_0 + 1 + \frac{1}{\alpha_{T^2} + \alpha_{\text{SPE}} + 2/(n_0 + 1)}.$$

Online Monitoring Procedure. The complete detection logic is summarized in Algorithm 1. At each time step t in the monitoring phase, the detector updates the Wasserstein tangent projection, projects deviations onto the pre-change reference eigenbasis, and triggers an alarm if either $T_{t,K}^2$ or SPE_t exceeds its calibrated threshold.

3.5. The ε -Isometry Guarantee.

The proposed online detector is efficient only if the sequence of tangent vectors can be accurately approximated. Therefore, we provide the following ε -isometry as a theoretical guarantee for this approximation, showing that the regularity of the optimal transport maps ensures a rapid spectral decay of the covariance operator. We begin by establishing the regularity of the pre-change process.

Assumption 3.12 (Regularity of transport maps). The pre-change process generates i.i.d. distributions $\{\mu_i\}$ on a convex, bounded domain $\mathcal{X} \subset \mathbb{R}^d$. Each μ_i admits a density ρ_i that is bounded and α -Hölder continuous (for some $\alpha > 0$). Let $T_t : \bar{\mu} \rightarrow \mu_t$ denote the unique optimal transport map from the reference barycenter $\bar{\mu}$ to μ_t , and define the tangent vector field $v_t(x) := T_t(x) - x$. By standard regularity results for Monge-Ampère equations (Brenier, 1991; Caffarelli, 1992; 2000), there exist deterministic constants $L, B < \infty$ such that almost surely:

$$\text{Lip}(v_t) \leq L \quad \text{and} \quad \|v_t\|_{L^\infty(\mathcal{X})} \leq B, \quad \text{for all } t. \quad (17)$$

Algorithm 1 IDD Online Change-Point Detector

Input : Reference stream $\{\mu_t\}_{t=1}^{n_0}$, truncation K , false alarm rates $\alpha_{T^2}, \alpha_{\text{SPE}}$.

Pre-change Calibration

Compute Wasserstein barycenter $\bar{\mu}$ of $\{\mu_t\}_{t=1}^{n_0}$ via Eq. (5).
for $t = 1, \dots, n_0$ **do**
 Solve OT between $\bar{\mu}$ and μ_t to get optimal plan π_t . Compute barycentric projection vector $v_t(x) = \int y \pi_t(dy | x) - x$.

end

Compute reference mean $\bar{v} = \frac{1}{n_0} \sum_{t=1}^{n_0} v_t$. Center the fields: $\tilde{v}_t = v_t - \bar{v}$ for $t = 1, \dots, n_0$.

Perform MFPCA on $\{\tilde{v}_t\}_{t=1}^{n_0}$ to obtain eigenpairs $\{(\hat{\phi}_m, \hat{\lambda}_m)\}_{m=1}^{n_0-1}$.
 Compute $\{T_{t,K}^2, \text{SPE}_t\}_{t=1}^{n_0}$ for the reference sample using Eq. (13) and Eq. (14). Set thresholds h_{T^2}, h_{SPE} as the $(1 - \alpha)$ empirical quantiles of $\{T_{t,K}^2, \text{SPE}_t\}_{t=1}^{n_0}$.

Online Monitoring

Initialize $\tau = \infty$. **for** $t = n_0 + 1, n_0 + 2, \dots$ **do**
 Receive μ_t , solve OT relative to $\bar{\mu}$, compute v_t . Center the new observation: $\Delta_t = v_t - \bar{v}$. Compute scores:
 $\hat{\xi}_{tm} = \langle \Delta_t, \hat{\phi}_m \rangle_H$ for $m = 1, \dots, K$.
 $T_{t,K}^2 = \sum_{m=1}^K \hat{\xi}_{tm}^2 / \hat{\lambda}_m$. $\text{SPE}_t = \|\Delta_t\|_H^2 - \sum_{m=1}^K \hat{\xi}_{tm}^2$.
if $T_{t,K}^2 > h_{T^2}$ **or** $\text{SPE}_t > h_{\text{SPE}}$ **then**
 | Flag t as a change point
end

end

This geometric regularity of the individual realizations implies the smoothness of the population covariance structure.

Proposition 3.13 (Lipschitz Continuity of the Covariance Kernel). *Under Assumption 3.12, let V be the random tangent field representing the pre-change process. The matrix-valued covariance kernel $\mathcal{K}(\mathbf{x}, \mathbf{y}) := \mathbb{E}[V(\mathbf{x})V(\mathbf{y})^\top]$ is Lipschitz continuous on $\mathcal{X} \times \mathcal{X}$. Specifically,*

$$\|\mathcal{K}(\mathbf{x}, \mathbf{y}) - \mathcal{K}(\mathbf{x}', \mathbf{y}')\|_F \leq C_{\mathcal{K}}(\|\mathbf{x} - \mathbf{x}'\| + \|\mathbf{y} - \mathbf{y}'\|),$$

where $C_{\mathcal{K}}$ depends on the Lipschitz constant L and the second moment of the process.

Proof. See Appendix B. \square

Leveraging the spectral theory of integral operators, the Lipschitz continuity of the kernel guarantees a specific decay rate for the eigenvalues of the covariance operator, allowing us to bound the truncated detector's approximation error.

Theorem 3.14 (ε -Isometry). *Let the domain $\mathcal{X} \subset \mathbb{R}^d$ and the covariance kernel \mathcal{K} satisfy the conditions of Proposition 3.13. Let $\{\lambda_m\}_{m \geq 1}$ be the eigenvalues of the covariance operator Γ in non-increasing order. Then, there exists*

a constant $A_{\mathcal{X}}$ depending only on the domain dimension d such that the tail sum of eigenvalues satisfies:

$$\sum_{m > K} \lambda_m \leq A_{\mathcal{X}} C_{\mathcal{K}} K^{-1/d}. \quad (18)$$

Consequently, to achieve an ε -isometry in mean square (i.e., relative reconstruction error $\leq \varepsilon^2$), the required number of principal components K scales as:

$$K \geq \left(\frac{A_{\mathcal{X}} C_{\mathcal{K}}}{\varepsilon^2 \text{tr}(\Gamma)} \right)^d \quad (19)$$

Proof. The proof extends the trace-bound arguments of Reade (1983) to d -dimensional Lipschitz domains via a dyadic partition argument. See Appendix B for the full derivation. \square

Remark 3.15. Theorem 3.14 provides the theoretical justification for our dimension reduction. It ensures that the Wasserstein tangent space can be accurately approximated by a low-dimensional Euclidean subspace with a truncation level K that grows polynomially with the desired precision.

4. Numerical Experiments

We evaluate the proposed IDD against baselines on synthetic batch streams (continuous and discrete) and two real-world case studies: (i) Acute Myeloid Leukemia (AML) detection (Aghaeepour et al., 2013), and (ii) Vaccine sentiment monitoring (Brambilla & Kharmale, 2022a). All implementation detail are provided in Appendix H, H, K L.

Baselines. We compare IDD against five representative baselines categorized by their underlying geometric representation. First, Euclidean summary baselines include Shewhart-type charts (Xie et al., 2021) (e.g., Hotelling's T^2 or attribute charts), which reduce distributions to simple moments like means or variances. Second, distance-based metric baselines utilize kernel or sketch-based discrepancies, specifically Scan-B (Li et al., 2019) (using Maximum Mean Discrepancy) and NEWMA (Keriven et al., 2020) (using Random Fourier Features). Third, manifold-value baselines explicitly model the space of probability measures but use alternative geometries (Ramsay & Silverman, 2005): Log-KDE monitors functional L^2 distances between log-density estimates, while F-CPD (Dubey & Müller, 2020) employs Fréchet variance on graph-based metrics.

Metrics. We evaluate performance via the standard false-alarm-delay trade-off. For a stopping time τ and change-point κ , we report the in-control average run length $\text{ARL}_0 = \mathbb{E}_{\infty}[\tau]$ (expected time to false alarm) and the detection delay $\text{ARL}_1 = \mathbb{E}_{\kappa}[\tau - \kappa | \tau > \kappa]$. All methods are calibrated to a fixed target ARL_0 to compare ARL_1 (smaller is better).

330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384

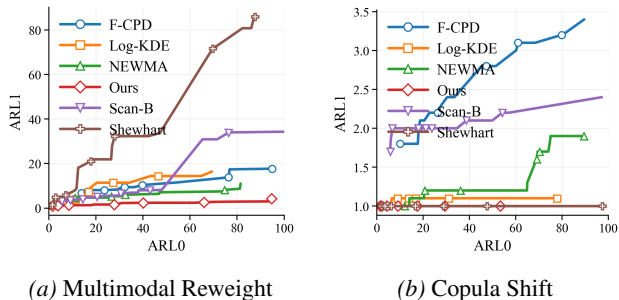


Figure 1. ARL_1 vs. ARL_0 comparison on synthetic continuous streams. IDD (red) shows superior detection speed.

4.1. Synthetic Experiments

We generate distribution-valued streams $\{\mu_t\}$ by pushing forward a reference measure $\bar{\mu}$ via convex potentials (continuous), where it is verified to yield optimal transport maps in Appendix G.3, and by simulating count and categorical processes (discrete).

Continuous Streams. We simulate d -dimensional streams ($d \in \{1, 5, 10, 50\}$) with batch sizes $N \in \{50, 100, 300\}$ under three shift scenarios: (1) Barycenter change: shifting the central mass; (2) Multimodal reweight: altering mixture weights of a multimodal distribution; and (3) Copula shift: altering variable dependencies while fixing marginals (see Appendix G for generation details). To ensure scalability in high-dimensional settings ($d \geq 10$), we accelerate computations using Sinkhorn iterations (Cuturi, 2013) and estimate barycenters via Conditional Normalizing Flows (Visentin & Cheridito, 2026) (see Appendix I).

Results: Results for Multimodal Reweight ($d = 10$) and Copula Shift ($d = 50$) are shown in Fig. 1a and Fig. 1b. IDD (red curve) consistently achieves the lowest ARL_1 at matched ARL_0 , demonstrating superior detection power. Notably, Log-KDE (orange square) remains robust and competitive across both dimensions. In contrast, F-CPD (blue circle) shows reduced sensitivity in the high-dimensional Copula shift ($d = 50$), which might be due to the inefficiency of its graph-based metric in sparse spaces. The Euclidean Shewhart baseline (brown) exhibits inconsistent performance: it fails catastrophically in the Multimodal setting (Fig. 1a) where means are preserved, but detects the covariance change in the Copula shift (Fig. 1b) effectively. Overall, IDD provides resilient performance across diverse scenarios, with full results in Appendix J.

Discrete Streams. We evaluate the framework on streams where observations lie on discrete supports. Details are provided in Appendix H. (1) Poisson Counts. For post-change, we introduce a Spike Injection shift, where a small fraction of counts are replaced by a fixed high-value outlier (k^*), simulating rare burst events. As shown in Fig. 2a, IDD

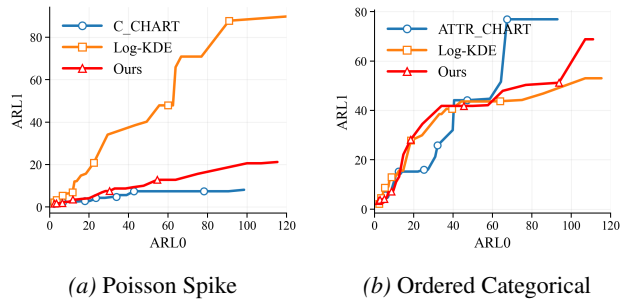


Figure 2. Trade-off curves for discrete streams ($N = 100$). IDD handles discrete geometry robustly.

Table 1. Gaussian translation: ARL_1 (mean \pm SE) at matched ARL_0 ; smaller is better. Best log-KDE bandwidth is reported.

d	σ	δ_1	Hotelling T^2	IDD (Ours)	Best log-KDE
1	0.5	0.1	1.8 ± 0.2 ($\uparrow 28.0\%$)	1.8 ± 0.2 ($\uparrow 28.0\%$)	2.5 ± 0.6
1	0.5	0.5	1.0 ± 0.0 ($\uparrow 95.1\%$)	1.0 ± 0.0 ($\uparrow 95.1\%$)	1.0 ± 0.0 ($\uparrow 95.1\%$)
1	1.0	0.1	4.2 ± 1.1 ($\uparrow 20.8\%$)	5.3 ± 1.6	10.8 ± 2.5
1	1.0	0.5	1.0 ± 0.0 ($\uparrow 9.1\%$)	1.0 ± 0.0 ($\uparrow 9.1\%$)	1.0 ± 0.0 ($\uparrow 9.1\%$)
1	2.0	0.1	16.7 ± 3.5	17.0 ± 3.4	16.4 ± 2.6
1	2.0	0.5	1.6 ± 0.3	1.5 ± 0.2 ($\uparrow 6.3\%$)	3.8 ± 1.3
2	0.5	0.1	1.5 ± 0.2	1.3 ± 0.2 ($\uparrow 13.3\%$)	2.3 ± 0.7
2	0.5	0.5	1.0 ± 0.0 ($\uparrow 96.4\%$)	1.0 ± 0.0 ($\uparrow 96.4\%$)	1.0 ± 0.0 ($\uparrow 96.4\%$)
2	1.0	0.1	2.7 ± 0.7	2.4 ± 0.6 ($\uparrow 11.1\%$)	5.7 ± 1.6
2	1.0	0.5	1.0 ± 0.0 ($\uparrow 9.1\%$)	1.0 ± 0.0 ($\uparrow 9.1\%$)	1.0 ± 0.0 ($\uparrow 9.1\%$)
2	2.0	0.1	11.3 ± 5.4	10.2 ± 5.3 ($\uparrow 9.7\%$)	15.4 ± 5.2
2	2.0	0.5	1.2 ± 0.1	1.0 ± 0.0 ($\uparrow 16.7\%$)	2.6 ± 0.6

significantly outperforms Log-KDE. This confirms that the kernel smoothing required by Log-KDE blurs distinct spikes on discrete grids, whereas IDD preserves signal fidelity by leveraging the intrinsic geometry of the full discrete distribution. (2) Ordered Categorical Drift. We simulate sequential survey-type data on an ordinal scale ($d = 6$ classes). The shift involves a Gradual Drift, where probability mass slowly migrates to adjacent higher classes. It can be shown that IDD outperforms Shewhart attribute charts (Fig. 2b). Unlike attribute charts, which treat classes as nominal, our method exploits the underlying ordinal metric. This allows it to detect the coherent flow of mass between adjacent ranks more efficiently than bin-wise monitoring.

Gaussian Translation Analysis. To rigorously quantify the geometric advantage, we compared IDD against Log-KDE on a pure Gaussian mean-shift $\mathcal{N}(m, \Sigma) \rightarrow \mathcal{N}(m + \delta/\sqrt{n}, \Sigma)$. We formalize this in **Theorem F.1** (Appendix F), which proves that KDE smoothing attenuates the signal in log-density space ($\|g_\Delta\|^2 \propto \delta^T (\Sigma + H)^{-1} \delta$), whereas the optimal transport signal remains invariant to smoothing. Empirical validation (Table 1, and Table 2 in Appendix) confirms this theory: IDD achieves up to a 95% reduction in detection delay compared to the best-tuned Log-KDE in high-variance settings.

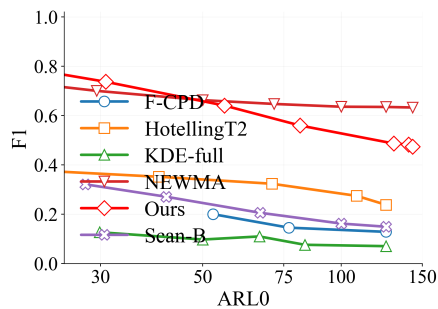


Figure 3. AML Detection on FlowCAP-II. IDD (red diamond) achieves the best F1-score (≈ 0.75) compared to baselines, balancing precision and speed.

4.2. Real-World Case Studies

1. Medical Diagnosis We apply our framework to the detection of Acute Myeloid Leukemia (AML) using the FlowCAP-II dataset (Aghaeepour et al., 2013). Flow cytometry data is naturally distribution-valued: each patient sample consists of thousands of single cells, where each cell is a vector of multidimensional fluorescence markers. We model these patient samples as a distribution-valued stream in \mathbb{R}^7 . The objective is to detect AML-positive subjects. We calibrate the detector on a reference set of healthy samples and monitor a test stream containing injected AML-positive patients (see Appendix K).

Results: As shown in Fig. 3, IDD achieves the highest F1-score (≈ 0.75) and fast detection ($ARL_1 \approx 1$). This indicates that our geometric-aware statistic successfully captures the intrinsic shape of the leukemic shift. In contrast, Hotelling’s T^2 yields low precision (< 0.4) because reducing distributions to simple moments discards critical sub-population structures. Meanwhile, Log-KDE suffers severe detection latency, confirming its struggle with high-dimensional density estimation.

2. Public Opinion Monitoring We analyze the semantic evolution of public discourse using the Reddit Vaccine Sentiment dataset (Brambilla & Kharmale, 2022a;b). The stream consists of daily batches of user comments from January to May 2021, embedded into a 20-dimensional space via SBERT and PCA (see Appendix L for data processing details). We focus on the critical early rollout phase, dividing the timeline into two distinct periods. Phase I serves as the pre-change calibration set (January 1 to February 26, 50 days), covering the relatively steady, initial rollout of Pfizer and Moderna vaccines. Phase II is designed to be the monitoring period (February 27 to May 5, 28 days), beginning with the Johnson & Johnson Emergency Use Authorization. This window contains high-impact shock events, most notably the J&J vaccine pause due to safety concerns and subsequent policy shifts.

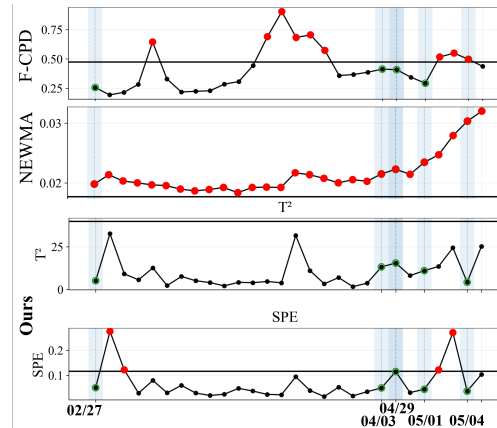


Figure 4. Reddit Sentiment Stream Monitoring. IDD alarms (red points) align with key news events (blue regions).

Results: Lacking ground truth labels for daily sentiment shifts, we evaluate performance qualitatively by correlating detection alarms with major news cycles. As shown in Fig. 4, IDD (*i.e.*, SPE statistics) generates distinct, sharp alarms that align closely with the J&J Pause on April 13 and the subsequent lifting of restrictions in late April. This alignment confirms that our method detects acute semantic shock. In contrast, the baselines fail to isolate these specific triggers. NEWMA results a monotonic upward drift, while F-CPD displays high variance unrelated to external events. This demonstrates the ability of IDD to filter background noise and highlight significant geometric shifts in high-dimensional semantic streams.

5. Discussion and Conclusion

We propose a geometry-aware online CPD framework that treats streaming batch data as a stochastic process on 2-Wasserstein space. By employing a reference-centered local linearization, our approach maps empirical measures into a tangent space. We provide theoretical results establishing an ϵ -isometry guarantee for this approximation, which enables the rigorous monitoring of distributional change that Euclidean summaries obscure. Empirically, our experiments on synthetic and real-world case studies validate that IDD achieves superior detection performance compared to moment-based and model-free baselines, particularly when monitoring complex distributional shifts.

We also identify the main limitations of our work: Despite acceleration via entropic regularization and Conditional Normalizing Flows, the computational cost of OT maps remains higher than simple Euclidean statistics. This poses challenges for extremely high-frequency streams. Future work will explore incremental barycenter updates, and investigate scalable OT approximations to mitigate computational latency and estimation error in high dimensions.

6. Impact statement

This paper presents work whose goal is to advance the field of machine learning. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

References

Aghaeepour, N., Finak, G., Consortium, F., Consortium, D., Hoos, H., Mosmann, T. R., Brinkman, R., Gottardo, R., and Scheuermann, R. H. Critical assessment of automated flow cytometry data analysis techniques. *Nature methods*, 10(3):228–238, 2013.

Aminikhanghahi, S. and Cook, D. J. A survey of methods for time series change point detection. *Knowledge and information systems*, 51(2):339–367, 2017. doi: 10.1007/s10115-016-0987-z. URL <https://doi.org/10.1007/s10115-016-0987-z>.

Aminikhanghahi, S., Wang, T., and Cook, D. J. Real-time change point detection with application to smart home time series data. *IEEE Transactions on Knowledge and Data Engineering*, 31(5):1010–1023, 2018.

Bakdi, A. and Kouadri, A. A new adaptive pca based thresholding scheme for fault detection in complex systems. *Chemometrics and Intelligent Laboratory Systems*, 162: 83–93, 2017.

Bhattacharya, A. and Bhattacharya, R. Statistics on riemannian manifolds: asymptotic distribution and curvature. *Proceedings of the American Mathematical Society*, 136(8):2959–2967, 2008.

Bhattacharya, R. and Patrangenaru, V. Large sample theory of intrinsic and extrinsic sample means on manifolds—ii. *The Annals of Statistics*, 33(3):1225–1259, 2005. doi: 10.1214/009053605000000093. URL <https://doi.org/10.1214/009053605000000093>.

Bigot, J., Gouet, R., Klein, T., and López, A. Geodesic pca in the wasserstein space by convex pca. *Annales de l’Institut Henri Poincaré (Probabilités et Statistiques)*, 53(1):1–26, 2017.

Bouchard, F., Mian, A., Zhou, J., Said, S., Ginolhac, G., and Berthoumieu, Y. Riemannian geometry for compound gaussian distributions: Application to recursive change detection. *Signal Processing*, 176:107716, 2020.

Brambilla, M. and Kharmale, K. COVID-19 Vaccine Discussions on Reddit with Sentiment, Stance, Topics, and Timing, 2022a. URL <https://doi.org/10.7910/DVN/XJTBQM>.

Brambilla, M. and Kharmale, K. Covid-19 vaccine discussions on reddit with sentiment. *Stance, Topics, and Timing*, 2022b.

Brenier, Y. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

Caffarelli, L. A. The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5(1):99–104, 1992.

Caffarelli, L. A. Monotonicity properties of optimal transportation and the fkg and related inequalities. *Communications in Mathematical Physics*, 214(3):547–563, 2000.

Chang, W.-C., Li, C.-L., Yang, Y., and Póczos, B. Kernel change-point detection with auxiliary deep generative models. In *International Conference on Learning Representations*, 2019. URL <https://arxiv.org/abs/1901.06077>.

Chernozhukov, V., Galichon, A., Hallin, M., and Henry, M. Monge–kantorovich depth, quantiles, ranks and signs. *Annals of Statistics*, 45(1):223–256, 2017.

Crosier, R. B. Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, 30(3):291–303, 1988.

Cuturi, M. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pp. 2292–2300, 2013.

Doob, J. L. *Stochastic Processes*. John Wiley & Sons, 1953.

Dubey, P. and Müller, H.-G. Fréchet change-point detection. *The Annals of Statistics*, 48(6):3312–3335, 2020.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. A kernel two-sample test. *The journal of machine learning research*, 13(1):723–773, 2012.

Hall, P. and Hosseini-Nasab, M. On properties of functional principal components analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(1):109–126, 2006.

Harchaoui, Z., Moulines, E., and Bach, F. R. Kernel change-point analysis. In *Advances in Neural Information Processing Systems*, 2008. URL <https://papers.nips.cc/paper/2008/hash/08b255a5d42b89b0585260b6f2360bdd-Abstract.html>.

Horváth, L., Kokoszka, P., and Wang, S. Monitoring for a change point in a sequence of distributions. *The Annals of Statistics*, 49(4):2271–2291, 2021.

- 495 Jackson, J. E. and Mudholkar, G. S. Control procedures for
496 residuals associated with principal component analysis.
497 *Technometrics*, 21(3):341–349, 1979.
- 498 Jiang, F., Zhu, C., and Shao, X. Two-sample and change-
499 point inference for non-euclidean valued time series. *Elec-
500 tronic Journal of Statistics*, 18(1):848–894, 2024.
- 502 Keriven, N., Garreau, D., and Poli, I. Newma: A new
503 method for scalable model-free online change-point detec-
504 tion. *IEEE Transactions on Signal Processing*, 68:3515–
505 3528, 2020. URL [https://ieeexplore.ieee.
506 org/document/9078835](https://ieeexplore.ieee.org/document/9078835).
- 508 Kuncheva, L. I. Change detection in streaming multivari-
509 ate data using likelihood detectors. *IEEE transactions
510 on knowledge and data engineering*, 25(5):1175–1180,
511 2011.
- 512 Li, S., Xie, Y., Dai, H., and Song, L. M-statistic
513 for kernel change-point detection. In *Ad-
514 vances in Neural Information Processing Sys-
515 tems*, pp. 3366–3374, 2015. URL [https://papers.nips.cc/paper/2015/hash/
516 eble78328c46506b46a4ac4a1e378b91-Abstract.
517 html](https://papers.nips.cc/paper/2015/hash/eb1e78328c46506b46a4ac4a1e378b91-Abstract.html).
- 520 Li, S., Xie, Y., Dai, H., and Song, L. Scan b-statistic for
521 kernel change-point detection. *Sequential Analysis*, 38
522 (4):503–544, 2019.
- 524 Lowry, C. A. and Montgomery, D. C. A review of multi-
525 variate control charts. *IIE transactions*, 27(6):800–810,
526 1995.
- 527 Mattingly, J. C., Miller, E., and Tran, D. Central limit theo-
528 rems for fréchet means on stratified spaces. *arXiv*, 2023.
529 URL <https://arxiv.org/abs/2311.09455>.
- 531 Montgomery, D. C. *Introduction to statistical quality con-
532 trol*. John wiley & sons, 2020.
- 534 Nelson, L. S. The shewhart control chart—tests for special
535 causes. *Journal of quality technology*, 16(4):237–239,
536 1984.
- 537 Page, E. S. Continuous inspection schemes. *Biometrika*, 41
538 (1/2):100–115, 1954.
- 540 Panaretos, V. M. and Zemel, Y. Statistical aspects
541 of wasserstein distances. *Annual review of statis-
542 tics and its application*, 6(1):405–431, 2019.
543 doi: 10.1146/annurev-statistics-030718-104938.
544 URL [https://doi.org/10.1146/
545 annurev-statistics-030718-104938](https://doi.org/10.1146/annurev-statistics-030718-104938).
- 546 Petersen, A. and Müller, H.-G. Functional data analysis
547 for density functions by transformation to a hilbert space.
548 *Annals of Statistics*, 44(1):183–218, 2016.
- 549 Ramsay, J. O. and Silverman, B. W. *Functional data analy-
sis*. Springer, 2005.
- Reade, J. Eigen-values of lipschitz kernels. In *Mathemati-
cal Proceedings of the Cambridge Philosophical Society*,
volume 93, pp. 135–140. Cambridge University Press,
1983.
- Reimers, N. and Gurevych, I. Sentence-bert: Sentence
embeddings using siamese bert-networks. *arXiv preprint
arXiv:1908.10084*, 2019.
- Shewhart, W. A. *Economic control of quality of manufac-
tured product*. D. Van Nostrand Company, Inc., New
York, 1931.
- Skubalska-Rafajłowicz, E. Random projections and
hotelling’s t2 statistics for change detection in high-
dimensional data streams. *International Journal of Ap-
plied Mathematics and Computer Science*, 23(2), 2013.
- Sommerfeld, M. and Munk, A. Inference for empirical
wasserstein distances on finite spaces. *Journal of the
Royal Statistical Society Series B: Statistical Methodol-
ogy*, 80(1):219–238, 2018.
- Székel, G. J. and Rizzo, M. L. Energy statistics: A class
of statistics based on distances. *Journal of statistical
planning and inference*, 143(8):1249–1272, 2013.
- Truong, C., Oudre, L., and Vayatis, N. Supervised ker-
nel change point detection with partial annotations. In
*ICASSP 2019-2019 IEEE International Conference on
Acoustics, Speech and Signal Processing (ICASSP)*, pp.
3147–3151. IEEE, 2019.
- Truong, C., Oudre, L., and Vayatis, N. Selective review of
offline change point detection methods. *Signal Process-
ing*, 167:107299, 2020.
- Villani, C. *Topics in Optimal Transportation*, volume 58 of
Graduate Studies in Mathematics. American Mathemat-
ical Society, 2003.
- Villani, C. *Optimal Transport: Old and New*, volume
338 of *Grundlehren der mathematischen Wissenschaften*.
Springer, Berlin, 2009. ISBN 978-3-540-71049-3.
- Villani, C. et al. *Optimal transport: old and new*, volume
338. Springer, 2008.
- Visentin, G. and Cheridito, P. Computing optimal transport
maps and wasserstein barycenters using conditional nor-
malizing flows. In *Forty-second International Conference
on Machine Learning*, 2026.
- Wang, H., Tang, M., Park, Y., and Priebe, C. E. Locality
statistics for anomaly detection in time series of graphs.
IEEE Transactions on Signal Processing, 62(3):703–717,
2013.

550 Wang, X., Borsoi, R. A., and Richard, C. Non-parametric
551 online change point detection on riemannian manifolds.
552 In *Proceedings of the 41st International Conference*
553 *on Machine Learning*, volume 235 of *Proceedings of*
554 *Machine Learning Research*, pp. 50143–50162. PMLR,
555 2024. URL [https://proceedings.mlr.press/](https://proceedings.mlr.press/v235/wang24f.html)
556 [v235/wang24f.html](https://proceedings.mlr.press/v235/wang24f.html).
557
558 Woodall, W. H. and Montgomery, D. C. Research issues
559 and ideas in statistical process control. *Journal of quality*
560 *technology*, 31(4):376–386, 1999.
561
562 Xie, L., Zou, S., Xie, Y., and Veeravalli, V. V. Sequential
563 (quickest) change detection: Classical results and new di-
564 rections. *IEEE Journal on Selected Areas in Information*
565 *Theory*, 2(2):494–514, 2021.
566
567 Zhang, Y., Zhu, C., and Shao, X. Change-point detection
568 for object-valued time series. *Journal of Business &*
569 *Economic Statistics*, (just-accepted):1–23, 2025.
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604

A. Assumptions

Lemma A.1 (Brenier map and unique W_2 -geodesic). *Assume $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ is absolutely continuous and the transport cost is $c(x, y) = \|x - y\|^2$. Then for any $\mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ there exists an optimal transport map $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ pushing μ_0 to μ_1 (i.e., $T_{\#}\mu_0 = \mu_1$), which is μ_0 -a.e. unique and of the form $T = \nabla\varphi$ for a convex potential φ . Moreover, the curve*

$$\mu_s := ((1-s)\text{Id} + sT)_{\#}\mu_0, \quad s \in [0, 1],$$

is the (unique) constant-speed W_2 -geodesic connecting μ_0 and μ_1 (Brenier, 1991; Villani et al., 2008).

B. Derivation of Radial Isometry and ε -isometry

B.1. Proof of Proposition 3.4

Proof. Under the quadratic cost, the squared 2-Wasserstein distance admits the Monge formulation

$$W_2^2(\bar{\mu}, \mu) = \inf_{T: T_{\#}\bar{\mu}=\mu} \int_{\mathbb{R}^d} \|x - T(x)\|^2 d\bar{\mu}(x).$$

Since $\bar{\mu}$ is absolutely continuous with respect to Lebesgue measure and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, Brenier's theorem guarantees the existence of an optimal transport map $T_{\bar{\mu}}^{\mu}$ that is $\bar{\mu}$ -a.e. unique and satisfies $(T_{\bar{\mu}}^{\mu})_{\#}\bar{\mu} = \mu$. Evaluating the Monge objective at this optimizer, we have

$$W_2^2(\bar{\mu}, \mu) = \int_{\mathbb{R}^d} \|x - T_{\bar{\mu}}^{\mu}(x)\|^2 d\bar{\mu}(x) = \int_{\mathbb{R}^d} \|T_{\bar{\mu}}^{\mu}(x) - x\|^2 d\bar{\mu}(x).$$

Defining $v(\mu) := T_{\bar{\mu}}^{\mu} - \text{Id}$ gives $W_2^2(\bar{\mu}, \mu) = \int \|v(\mu)(x)\|^2 d\bar{\mu}(x) = \|v(\mu)\|_{L^2(\bar{\mu})}^2$, completing the proof. \square

B.2. Proof of Proposition 3.6

Proof. Let $(X, Y) \sim \pi$ and define $T^{\pi}(X) := \mathbb{E}[Y | X]$. By the conditional-variance identity,

$$\mathbb{E}[\|Y - T^{\pi}(X)\|^2] = \mathbb{E}[\|Y\|^2] - \mathbb{E}[\|T^{\pi}(X)\|^2].$$

Expanding the squared cost and using $\mathbb{E}[Y^{\top} T^{\pi}(X)] = \mathbb{E}[\mathbb{E}[Y | X]^{\top} T^{\pi}(X)] = \mathbb{E}[\|T^{\pi}(X)\|^2]$, we obtain

$$\begin{aligned} \mathbb{E}[\|X - Y\|^2] &= \mathbb{E}[\|X\|^2] - 2\mathbb{E}[X^{\top} Y] + \mathbb{E}[\|Y\|^2] \\ &= \mathbb{E}[\|X\|^2] - 2\mathbb{E}[X^{\top} T^{\pi}(X)] + \mathbb{E}[\|T^{\pi}(X)\|^2] + \mathbb{E}[\|Y - T^{\pi}(X)\|^2] \\ &= \mathbb{E}[\|X - T^{\pi}(X)\|^2] + \mathbb{E}[\|Y - T^{\pi}(X)\|^2], \end{aligned}$$

which is exactly Eq. (9) after rewriting expectations as integrals with respect to μ and π . The inequality Eq. (10) follows since the second term is nonnegative. Finally, for an optimal coupling π^* , $\mathbb{E}[\|X - Y\|^2] = W_2^2(\mu, \nu)$ and equality in Eq. (10) holds if and only if $\mathbb{E}[\|Y - T^{\pi^*}(X)\|^2] = 0$, i.e., $Y = T^{\pi^*}(X)$ π^* -a.s., equivalently $\text{Var}_{\pi^*}(Y | X) = 0$ μ -a.e. \square

B.3. Proof of Proposition 3.13

Proof. We bound the Frobenius norm of the kernel difference. By the triangle inequality:

$$\begin{aligned} |\mathcal{K}(\mathbf{x}, \mathbf{y}) - \mathcal{K}(\mathbf{x}', \mathbf{y}')|_F &= |\mathbb{E}[V(\mathbf{x})V(\mathbf{y})^{\top} - V(\mathbf{x}')V(\mathbf{y}')^{\top}]|_F \\ &\leq \mathbb{E}|V(\mathbf{x})V(\mathbf{y})^{\top} - V(\mathbf{x}')V(\mathbf{y}')^{\top}|_F + \mathbb{E}|V(\mathbf{x}')V(\mathbf{y})^{\top} - V(\mathbf{x}')V(\mathbf{y}')^{\top}|_F. \end{aligned} \quad (20)$$

Applying the Lipschitz property $\|V(\mathbf{x}) - V(\mathbf{x}')\| \leq L\|\mathbf{x} - \mathbf{x}'\|$ (Assumption 3.12) and the Cauchy-Schwarz inequality:

$$\mathbb{E}\|(V(\mathbf{x}) - V(\mathbf{x}'))V(\mathbf{y})^{\top}\|_F \leq \sqrt{\mathbb{E}\|V(\mathbf{x}) - V(\mathbf{x}')\|^2} \sqrt{\mathbb{E}\|V(\mathbf{y})\|^2} \leq L\|\mathbf{x} - \mathbf{x}'\|B_2,$$

where $B_2 := \sup_{\mathbf{z} \in \mathcal{X}} (\mathbb{E}\|V(\mathbf{z})\|^2)^{1/2}$ is the maximum root-mean-square norm. Applying the same bound to the second term yields:

$$\|\mathcal{K}(\mathbf{x}, \mathbf{y}) - \mathcal{K}(\mathbf{x}', \mathbf{y}')\|_F \leq LB_2\|\mathbf{x} - \mathbf{x}'\| + LB_2\|\mathbf{y} - \mathbf{y}'\|.$$

Thus, \mathcal{K} is Lipschitz with constant $C_{\mathcal{K}} = LB_2$. \square

B.4. Proof of Theorem 3.14

Proof. The proof proceeds in three steps, generalizing the 1D results of [Reade \(1983\)](#). **Step 1: Partition and Projector.** Fix $M \in \mathbb{N}$ and set $N = M^d$. Partition the domain \mathcal{X} into N measurable sets $\{\mathcal{X}_j\}_{j=1}^N$ with diameter proportional to $M^{-1} \asymp N^{-1/d}$. Let R_N be the kernel of the orthogonal projector onto piecewise constant functions on this partition. **Step 2: Trace Bound.** Let S_N be the rank- N approximation of the kernel operator Γ induced by this projection. Since Γ is positive definite, the tail sum of eigenvalues is bounded by the trace of the residual operator:

$$\sum_{m>N} \lambda_m \leq \text{tr}(\Gamma - S_N) = \int_{\mathcal{X}} (\mathcal{K}(\mathbf{x}, \mathbf{x}) - S_N(\mathbf{x}, \mathbf{x})) d\mathbf{x}.$$

Step 3: Lipschitz Estimate. Using the Lipschitz continuity of \mathcal{K} ([Proposition 3.13](#)), the pointwise approximation error of the piecewise constant projector is bounded by the Lipschitz constant $C_{\mathcal{K}}$ times the diameter of the partition cells:

$$|\mathcal{K}(\mathbf{x}, \mathbf{x}) - S_N(\mathbf{x}, \mathbf{x})| \lesssim C_{\mathcal{K}} \cdot \text{diam}(\mathcal{X}_j) \asymp C_{\mathcal{K}} N^{-1/d}.$$

Integrating this error over the bounded domain \mathcal{X} yields the tail bound:

$$\sum_{m>N} \lambda_m \leq A_{\mathcal{X}} C_{\mathcal{K}} N^{-1/d}.$$

Setting the truncation level $K = N$ completes the proof. □

C. Derivation of Asymptotic Distributions

Proof of Proposition 3.9. Let V denote the generic random field generating the pre-change process, with population covariance Γ and eigenpairs (λ_m, ϕ_m) . By the Karhunen-Loève theorem, the random field V admits the expansion $V = \sum_{m=1}^{\infty} \xi_m \phi_m$, where the scores $\xi_m = \langle V, \phi_m \rangle_H$ are independent Gaussian variables $\xi_m \sim \mathcal{N}(0, \lambda_m)$. Consider the empirical covariance \hat{C} estimated from n_0 samples, with eigenpairs $(\hat{\lambda}_m, \hat{\phi}_m)$. For a new independent observation V^* , the computed score on the m -th component is $\hat{\xi}_m = \langle V^*, \hat{\phi}_m \rangle_H$. Under the conditions of [Hall & Hosseini-Nasab \(2006\)](#), as $n_0 \rightarrow \infty$, we have consistent estimation:

$$\|\hat{\phi}_m - \phi_m\|_H = O_p(n_0^{-1/2}) \quad \text{and} \quad |\hat{\lambda}_m - \lambda_m| = O_p(n_0^{-1/2}).$$

Consequently, $\hat{\xi}_m^2 / \hat{\lambda}_m \xrightarrow{p} \xi_m^2 / \lambda_m$. Since $\xi_m / \sqrt{\lambda_m} \sim \mathcal{N}(0, 1)$, it follows that $\xi_m^2 / \lambda_m \sim \chi_1^2$. Summing over $m = 1, \dots, K$ yields:

$$T_{t,K}^2 = \sum_{m=1}^K \frac{\hat{\xi}_m^2}{\hat{\lambda}_m} \xrightarrow{d} \sum_{m=1}^K Z_m^2 \sim \chi_K^2,$$

which concludes the proof. □

D. Details on Threshold Calibration and ARL Control

In this appendix, we provide the formal justification for the empirical quantile calibration strategy used in the main text. We show that controlling the marginal false alarm rates of the monitoring statistics via order statistics is sufficient to lower-bound the global ARL_0 .

Theorem D.1 (Sequential false-alarm control under fixed thresholds). *Fix truncation K and thresholds h_{T^2}, h_{SPE} . Under the no-change law \mathbb{P}_{∞} , assume the monitoring-phase pairs $\{(T_{t,K}^2, \text{SPE}_t)\}_{t \geq n_0+1}$ are i.i.d. and independent of the calibration phase segment used to construct $(h_{T^2}, h_{\text{SPE}})$. Let the stopping time*

$$\tau := \inf \left\{ t \geq n_0 + 1 : T_{t,K}^2 > h_{T^2} \text{ or } \text{SPE}_t > h_{\text{SPE}} \right\}.$$

Then $\tau - (n_0 + 1)$ has a geometric distribution with success probability (e.g., the classical run-length calculation for Shewhart-type charts; [Shewhart, 1931](#); [Montgomery, 2020](#))

$$p_{\infty} := \mathbb{P}_{\infty} \left(T_{1,K}^2 > h_{T^2} \text{ or } \text{SPE}_1 > h_{\text{SPE}} \right),$$

and hence $\text{ARL}_0 = \mathbb{E}_\infty[\tau] = n_0 + 1 + 1/p_\infty$. Moreover, if $\mathbb{P}_\infty(T_{1,K}^2 > h_{T^2}) \leq \alpha_{T^2}$ and $\mathbb{P}_\infty(\text{SPE}_1 > h_{\text{SPE}}) \leq \alpha_{\text{SPE}}$, then by the union bound $p_\infty \leq \alpha_{T^2} + \alpha_{\text{SPE}}$ and therefore

$$\text{ARL}_0 \geq n_0 + 1 + \frac{1}{\alpha_{T^2} + \alpha_{\text{SPE}}}.$$

Corollary D.2 (Empirical-quantile calibration implies ARL control). *Assume the setting of Theorem 3.10. Suppose that under \mathbb{P}_∞ the marginal distributions of $T_{t,K}^2$ and SPE_t are continuous, and let $\{T_{t,K}^2\}_{t=1}^{n_0}$ and $\{\text{SPE}_t\}_{t=1}^{n_0}$ be the samples in the calibration phase. Define thresholds as order statistics*

$$h_{T^2} := T_{(k_{T^2})}^2, \quad h_{\text{SPE}} := \text{SPE}_{(k_{\text{SPE}})},$$

where $T_{(1)}^2 \leq \dots \leq T_{(n_0)}^2$ and $\text{SPE}_{(1)} \leq \dots \leq \text{SPE}_{(n_0)}$, with $k_{T^2} = \lceil (1 - \alpha_{T^2})n_0 \rceil$ and $k_{\text{SPE}} = \lceil (1 - \alpha_{\text{SPE}})n_0 \rceil$. Then, for an independent monitoring phase observation,

$$\begin{aligned} \mathbb{P}_\infty(T_{1,K}^2 > h_{T^2}) &= \frac{n_0 + 1 - k_{T^2}}{n_0 + 1} \leq \alpha_{T^2} + \frac{1}{n_0 + 1}, \\ \mathbb{P}_\infty(\text{SPE}_1 > h_{\text{SPE}}) &= \frac{n_0 + 1 - k_{\text{SPE}}}{n_0 + 1} \leq \alpha_{\text{SPE}} + \frac{1}{n_0 + 1}. \end{aligned}$$

Consequently,

$$\begin{aligned} p_\infty &\leq \alpha_{T^2} + \alpha_{\text{SPE}} + \frac{2}{n_0 + 1}, \\ \text{ARL}_0 &\geq n_0 + 1 + \frac{1}{\alpha_{T^2} + \alpha_{\text{SPE}} + 2/(n_0 + 1)}. \end{aligned}$$

E. Construction of Tangent Space Detectors

Pre-change reference segment and covariance operator. In online CPD, we assume access to a *pre-change* reference segment of length n_0 (used only for calibration), which provides tangent fields $\{v_t\}_{t=1}^{n_0}$ in the common Hilbert space $H := L^2(\bar{\mu}; \mathbb{R}^d)$ with inner product $\langle f, g \rangle_H := \int_{\mathcal{X}} f(x)^\top g(x) d\bar{\mu}(x)$. Define the sample mean and centered fields

$$\bar{v} := \frac{1}{n_0} \sum_{t=1}^{n_0} v_t, \quad \tilde{v}_t := v_t - \bar{v}.$$

The empirical covariance operator $\hat{C} : H \rightarrow H$ is

$$(\hat{C}f)(\cdot) = \frac{1}{n_0 - 1} \sum_{t=1}^{n_0} \langle f, \tilde{v}_t \rangle_H \tilde{v}_t(\cdot) = \frac{1}{n_0 - 1} \sum_{t=1}^{n_0} (\tilde{v}_t \otimes \tilde{v}_t) f, \quad (21)$$

where $(a \otimes b)f := \langle f, b \rangle_H a$. Then \hat{C} is self-adjoint and positive semidefinite, and $\text{rank}(\hat{C}) \leq n_0 - 1$.

Hotelling- T^2 statistic via the Moore-Penrose pseudoinverse. Because \hat{C} is finite-rank on the infinite-dimensional space H , it is not invertible. We therefore use the Moore-Penrose pseudoinverse $\hat{C}^\dagger : H \rightarrow H$. Let $\text{ran}(\hat{C})$ and $\text{ker}(\hat{C})$ denote the range and kernel of \hat{C} . Since \hat{C} is self-adjoint, H decomposes as

$$H = \text{ran}(\hat{C}) \oplus \text{ker}(\hat{C}), \quad \text{ker}(\hat{C}) = \text{ran}(\hat{C})^\perp.$$

The pseudoinverse \hat{C}^\dagger is the unique self-adjoint operator satisfying

$$\hat{C} \hat{C}^\dagger \hat{C} = \hat{C}, \quad \hat{C}^\dagger \hat{C} \hat{C}^\dagger = \hat{C}^\dagger, \quad (\hat{C} \hat{C}^\dagger)^* = \hat{C} \hat{C}^\dagger, \quad (\hat{C}^\dagger \hat{C})^* = \hat{C}^\dagger \hat{C}.$$

For an incoming (possibly post-change) tangent field $v_t \in H$, define the centered field

$$\Delta_t := v_t - \bar{v} \in H.$$

We define the functional Hotelling's T^2 statistic-based detector

$$T_t^2 := \langle \hat{C}^\dagger \Delta_t, \Delta_t \rangle_H = \|(\hat{C}^\dagger)^{1/2} \Delta_t\|_H^2, \quad (22)$$

which measures deviation along the dominant pre-change variability directions encoded by \hat{C} .

Spectral form of the pseudoinverse and the computable T^2 statistic. Let $\{(\widehat{\lambda}_m, \widehat{\phi}_m)\}_{m \geq 1}$ be the eigenpairs of \widehat{C} , with $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \dots \geq 0$ and $\{\widehat{\phi}_m\}$ orthonormal in H . Since \widehat{C} has rank at most $n_0 - 1$, we have $\widehat{\lambda}_m = 0$ for all $m \geq n_0$. On $\text{ran}(\widehat{C})$, the pseudoinverse acts by inverting the nonzero eigenvalues, hence

$$\widehat{C}^\dagger = \sum_{m: \widehat{\lambda}_m > 0} \widehat{\lambda}_m^{-1} (\widehat{\phi}_m \otimes \widehat{\phi}_m), \quad (23)$$

with $\widehat{C}^\dagger f = 0$ for $f \in \ker(\widehat{C})$. Define the empirical scores for Δ_t ,

$$\widehat{\xi}_{tm} := \langle \Delta_t, \widehat{\phi}_m \rangle_H.$$

Substituting Eq. (23) into Eq. (22) yields

$$T_t^2 = \sum_{m: \widehat{\lambda}_m > 0} \frac{\widehat{\xi}_{tm}^2}{\widehat{\lambda}_m}.$$

In practice we retain the leading K components (regularization / truncation), obtaining

$$T_{t,K}^2 := \sum_{m=1}^K \frac{\widehat{\xi}_{tm}^2}{\widehat{\lambda}_m}. \quad (24)$$

Sequential false-alarm control under i.i.d. monitoring statistics

Proof of Theorem 3.10. Let $A_t := \{T_{t,K}^2 > h_{T^2} \text{ or } \text{SPE}_t > h_{\text{SPE}}\}$ for $t \geq n_0 + 1$. Under \mathbb{P}_∞ , by the assumed i.i.d. structure and the fact that thresholds are fixed in monitoring phase, the indicators $\{\mathbf{1}(A_t)\}_{t \geq n_0+1}$ are i.i.d. Bernoulli with success probability $p_\infty = \mathbb{P}_\infty(A_{n_0+1})$. Therefore, $\tau - (n_0 + 1)$ is geometric with mean $1/p_\infty$, giving $\mathbb{E}_\infty[\tau] = n_0 + 1 + 1/p_\infty$. The union-bound inequality follows from $p_\infty = \mathbb{P}_\infty(A_{n_0+1}) \leq \mathbb{P}_\infty(T_{1,K}^2 > h_{T^2}) + \mathbb{P}_\infty(\text{SPE}_1 > h_{\text{SPE}}) \leq \alpha_{T^2} + \alpha_{\text{SPE}}$. \square

Proof of Corollary D.2. We prove the statement for $T_{t,K}^2$; the argument for SPE_t is identical. Let Z_1, \dots, Z_{n_0} be the sample of calibration phase and let Z^* be an independent monitoring phase draw, all i.i.d. from a continuous distribution. Let $Z_{(1)} \leq \dots \leq Z_{(n_0)}$ denote the order statistics and set $h := Z_{(k)}$. By exchangeability of $(Z_1, \dots, Z_{n_0}, Z^*)$ and continuity (no ties a.s.), the rank of Z^* among these $n_0 + 1$ values is uniform on $\{1, \dots, n_0 + 1\}$. Therefore,

$$\mathbb{P}(Z^* > h) = \mathbb{P}(\text{rank}(Z^*) \geq k + 1) = \frac{n_0 + 1 - k}{n_0 + 1}.$$

With $k = \lceil (1 - \alpha)n_0 \rceil$, we have $n_0 - k \leq \alpha n_0$, hence $(n_0 + 1 - k)/(n_0 + 1) \leq \alpha + 1/(n_0 + 1)$. Applying this bound to both charts and using the union bound gives $p_\infty \leq \alpha_{T^2} + \alpha_{\text{SPE}} + 2/(n_0 + 1)$. The ARL bound follows from Theorem 3.10. \square

Residual-energy statistic (SPE/Q). Let \widehat{P}_K be the orthogonal projector onto $\text{span}\{\widehat{\phi}_1, \dots, \widehat{\phi}_K\}$. The residual-energy detector is

$$\text{SPE}_t := \|(I - \widehat{P}_K)\Delta_t\|_H^2 = \sum_{m > K} \widehat{\xi}_{tm}^2. \quad (25)$$

The pair $\{T_{t,K}^2, \text{SPE}_t\}$ forms a two-statistic online CPD rule: an alarm is raised when either statistic exceeds its pre-change calibrated threshold.

F. Theoretical Analysis of Gaussian Translation

In this appendix, we analyze the theoretical signal-to-noise ratio (effect size) of the proposed geometric framework compared to the functional log-density baseline. We consider a controlled Gaussian mean-shift scenario to derive explicit expressions for the displacement signal in both the Optimal Transport tangent space and the L^2 log-density space.

Theorem F.1 (Effect Size Comparison). Consider a reference distribution $\bar{\mu} = \mathcal{N}(m, \Sigma)$ and a perturbed distribution $\mu_{\Delta_n} = \mathcal{N}(m + \Delta_n, \Sigma)$ representing a local mean-shift with $\Delta_n = \delta/\sqrt{n}$. As $n \rightarrow \infty$, the squared norms of the shift in the OT tangent space (H_{OT}) and the log-density space (H_{log}) satisfy:

$$\|u_{\Delta_n}\|_{H_{\text{OT}}}^2 = \frac{1}{n} \|\delta\|^2, \quad (26)$$

$$\|g_{\Delta_n} - g_0\|_{H_{\text{log}}}^2 = \frac{1}{n} \delta^\top \Sigma^{-1} \delta + O\left(\frac{1}{n^2}\right). \quad (27)$$

Consequently, the ratio of the effect sizes is determined by the covariance structure:

$$\frac{\|g_{\Delta_n} - g_0\|_{H_{\text{log}}}^2}{\|u_{\Delta_n}\|_{H_{\text{OT}}}^2} = \frac{\delta^\top \Sigma^{-1} \delta}{\|\delta\|^2} + O\left(\frac{1}{n}\right). \quad (28)$$

Proof. We first derive the norm in the Optimal Transport tangent space. For a pure Gaussian translation from $\mathcal{N}(m, \Sigma)$ to $\mathcal{N}(m + \Delta, \Sigma)$, the unique optimal transport map is the identity shift $T(x) = x + \Delta$. The corresponding tangent vector field is the constant map $u_\Delta(x) \equiv \Delta$. The norm in the tangent space $H_{\text{OT}} = L^2(\bar{\mu}; \mathbb{R}^d)$ is calculated directly by integrating the squared Euclidean magnitude of the displacement against the reference measure:

$$\|u_\Delta\|_{H_{\text{OT}}}^2 = \int_{\mathbb{R}^d} \|u_\Delta(x)\|^2 d\bar{\mu}(x) = \|\Delta\|^2.$$

Substituting the local alternative $\Delta_n = \delta/\sqrt{n}$, we obtain $\|u_{\Delta_n}\|_{H_{\text{OT}}}^2 = \frac{1}{n} \|\delta\|^2$. Notably, this norm depends strictly on the Euclidean magnitude of the shift and is independent of the distribution's covariance Σ .

Next, we analyze the norm in the log-density space $H_{\text{log}} = L^2(\bar{\mu}; \mathbb{R})$. Let g_0 and g_Δ denote the log-densities of the reference $\bar{\mu}$ and the shifted measure μ_Δ , respectively. Defining the centered random variable $Y = X - m$, the difference in log-densities simplifies to:

$$g_\Delta(X) - g_0(X) = (\Sigma^{-1} \Delta)^\top Y - \frac{1}{2} \Delta^\top \Sigma^{-1} \Delta.$$

The squared norm in H_{log} corresponds to the second moment of this difference under the reference measure $\bar{\mu}$. Letting $a = \Sigma^{-1} \Delta$ and $c = \frac{1}{2} \Delta^\top \Sigma^{-1} \Delta$, and noting that $Y \sim \mathcal{N}(0, \Sigma)$, we have $\mathbb{E}[a^\top Y] = 0$ and $\text{Var}(a^\top Y) = a^\top \Sigma a$. Thus:

$$\begin{aligned} \|g_\Delta - g_0\|_{H_{\text{log}}}^2 &= \text{Var}(a^\top Y) + (\mathbb{E}[a^\top Y - c])^2 \\ &= a^\top \Sigma a + c^2 \\ &= (\Delta^\top \Sigma^{-1}) \Sigma (\Sigma^{-1} \Delta) + \left(\frac{1}{2} \Delta^\top \Sigma^{-1} \Delta\right)^2 \\ &= \Delta^\top \Sigma^{-1} \Delta + O(\|\Delta\|^4). \end{aligned}$$

Substituting $\Delta_n = \delta/\sqrt{n}$, the leading term is $\frac{1}{n} \delta^\top \Sigma^{-1} \delta$, which yields Eq. equation 27. \square

Remark F.2 (Effect of KDE Smoothing). The result above highlights a critical practical difference. The Log-KDE baseline typically estimates densities using a kernel bandwidth matrix H , which inflates the effective covariance to $\Sigma_{\text{eff}} \approx \Sigma + H$. Substituting Σ_{eff} into Eq. equation 27 shows that the signal strength in the log-density space is attenuated by smoothing:

$$\delta^\top (\Sigma + H)^{-1} \delta < \delta^\top \Sigma^{-1} \delta.$$

In high-variance or high-dimensional regimes where large bandwidths are required for stability, this attenuation is significant. In contrast, the OT signal (Eq. equation 26) remains $\|\delta\|^2/n$, invariant to both the intrinsic covariance Σ and any smoothing parameters.

F.1. Additional Empirical Results: Gaussian Translation

Table 2 presents the detailed ARL_1 comparison for Gaussian mean shifts across dimensions $d \in \{1, 2, 5\}$, variances σ , and shift magnitudes δ . IDD consistently outperforms Log-KDE, particularly in high-variance regimes ($\sigma = 2.0$) where Log-KDE requires larger bandwidths to stabilize.

Table 2. Gaussian translation detection delay (ARL₁) at matched ARL₀. The Ours column is shaded; best (smallest) per row is bolded.

d	σ	δ_1	Hotelling T^2	Ours	log-KDE ($h=0.5$)	log-KDE ($h=1$)	log-KDE ($h=1.5$)	log-KDE ($h=auto$)
1	0.5	0.1	1.8 ± 0.2 (↑28.0%)	1.8 ± 0.2 (↑28.0%)	4.7 ± 2.2	2.5 ± 0.6	2.6 ± 0.5	18.3 ± 6.8
1	0.5	0.5	1.0 ± 0.0 (↑95.1%)	1.0 ± 0.0 (↑95.1%)	1.0 ± 0.0 (↑95.1%)	1.0 ± 0.0 (↑95.1%)	1.0 ± 0.0 (↑95.1%)	1.0 ± 0.0 (↑95.1%)
1	1.0	0.1	4.2 ± 1.1 (↑20.8%)	5.3 ± 1.6	10.8 ± 2.5	19.0 ± 10.8	17.4 ± 9.7	21.5 ± 6.8
1	1.0	0.5	1.0 ± 0.0 (↑9.1%)	1.0 ± 0.0 (↑9.1%)	1.0 ± 0.0 (↑9.1%)	1.1 ± 0.1	1.1 ± 0.1	2.0 ± 0.7
1	2.0	0.1	16.7 ± 3.5	17.0 ± 3.4	16.4 ± 2.6	21.5 ± 6.9	19.9 ± 4.8	17.1 ± 2.4
1	2.0	0.5	1.6 ± 0.3	1.5 ± 0.2 (↑6.3%)	7.8 ± 3.4	6.6 ± 3.6	3.8 ± 1.3	12.9 ± 4.0
2	0.5	0.1	1.5 ± 0.2	1.3 ± 0.2 (↑13.3%)	2.3 ± 0.7	5.6 ± 3.5	9.4 ± 7.6	12.3 ± 3.0
2	0.5	0.5	1.0 ± 0.0 (↑96.4%)	1.0 ± 0.0 (↑96.4%)	1.0 ± 0.0 (↑96.4%)	1.0 ± 0.0 (↑96.4%)	1.0 ± 0.0 (↑96.4%)	1.0 ± 0.0 (↑96.4%)
2	1.0	0.1	2.7 ± 0.7	2.4 ± 0.6 (↑11.1%)	32.5 ± 19.3	5.7 ± 1.6	6.3 ± 1.6	16.1 ± 3.9
2	1.0	0.5	1.0 ± 0.0 (↑9.1%)	1.0 ± 0.0 (↑9.1%)	1.1 ± 0.1	1.0 ± 0.0 (↑9.1%)	1.0 ± 0.0 (↑9.1%)	1.5 ± 0.3
2	2.0	0.1	11.3 ± 5.4	10.2 ± 5.3 (↑9.7%)	16.9 ± 5.3	18.9 ± 7.9	23.4 ± 10.0	15.4 ± 5.2
2	2.0	0.5	1.2 ± 0.1	1.0 ± 0.0 (↑16.7%)	6.9 ± 1.5	3.3 ± 0.7	2.6 ± 0.6	7.2 ± 1.5
5	0.5	0.1	1.0 ± 0.0 (↑16.7%)	1.9 ± 0.5	6.0 ± 1.5	1.2 ± 0.2	1.7 ± 0.4	12.6 ± 3.5
5	0.5	0.5	1.0 ± 0.0 (↑96.3%)	1.0 ± 0.0 (↑96.3%)	1.0 ± 0.0 (↑96.3%)	1.0 ± 0.0 (↑96.3%)	1.0 ± 0.0 (↑96.3%)	1.0 ± 0.0 (↑96.3%)
5	1.0	0.1	2.8 ± 0.7 (↑42.9%)	4.9 ± 1.7	14.7 ± 4.3	11.9 ± 4.3	9.4 ± 2.3	11.1 ± 3.2
5	1.0	0.5	1.0 ± 0.0 (↑93.2%)	1.0 ± 0.0 (↑93.2%)	1.0 ± 0.0 (↑93.2%)	1.0 ± 0.0 (↑93.2%)	1.0 ± 0.0 (↑93.2%)	1.0 ± 0.0 (↑93.2%)
5	2.0	0.1	6.9 ± 1.9 (↑37.3%)	11.0 ± 3.2	26.5 ± 4.2	22.1 ± 4.7	19.6 ± 4.4	16.0 ± 4.0
5	2.0	0.5	1.0 ± 0.0 (↑9.1%)	1.1 ± 0.1	21.7 ± 5.8	4.4 ± 1.4	4.0 ± 0.8	8.8 ± 2.7

G. Details on Synthetic Continuous Stream Generation

This appendix provides the full specification of the synthetic distribution-valued streams used in Section 4.1. We detail: (i) the pre-change reference distribution, (ii) the family of convex deformations used to generate stochastic variability within a stationary regime, (iii) the post-change scenarios, and (iv) a theoretical verification that the generated maps satisfy cyclic monotonicity, thereby coinciding with the quadratic-cost optimal transport (OT) maps.

G.1. Pre-change Reference Distribution

Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact convex domain (in our experiments, $\mathcal{X} = [0, 1]^d$). We first sample a *reference distribution* $\bar{\mu} \in \mathcal{P}_2(\mathcal{X})$ from a multimodal parametric family. From this distribution, we draw a fixed *base sample* $X_{\text{base}}^{(0)} = \{x_i\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \bar{\mu}$. We utilize a K -component mixture of product-Beta distributions, parameterized by mixture weights $\pi \in \Delta^{K-1}$ (where Δ^{K-1} denotes the simplex) and component parameters θ_k . For the experiments, we set $K = 4$.

G.2. Stochastic Variability via Convex Deformations

To model benign variability within a stationary regime, at each time step t , we generate a random deformation map $T_t : \mathcal{X} \rightarrow \mathbb{R}^d$. The observed batch data is generated as the pushforward of the base sample: $X_t = T_t(X_{\text{base}})$. The resulting empirical measure is:

$$\mu_t = \frac{1}{N} \sum_{i=1}^N \delta_{(X_t)_i}.$$

Convex Potential and Transport Map. To ensure the deformation corresponds to an optimal transport map, we construct T_t as the gradient of a strictly convex potential Φ_t . We define this potential as a perturbation of the identity potential $\frac{1}{2}\|x\|^2$. Fix integers $J \geq 1$ and a smoothness parameter $\beta > 0$. At each time t , we draw random directions $a_{t,j} \in \mathbb{S}^{d-1}$, weights $w_{t,j} > 0$ (normalized such that $\sum_j w_{t,j} = 1$), and offsets $c_{t,j} \in \mathbb{R}$. Let $\zeta_\beta(z) = \beta^{-1} \log(1 + e^{\beta z})$ be the softplus function and $\sigma_\beta(z) = (1 + e^{-\beta z})^{-1}$ be the sigmoid function. We define a convex perturbation potential ψ_t as:

$$\psi_t(x) := \frac{\varepsilon_t}{2} \sum_{j=1}^J w_{t,j} (\zeta_\beta(\langle a_{t,j}, x \rangle - c_{t,j}))^2.$$

The full transport potential is $\Phi_t(x) := \frac{1}{2}\|x\|^2 + \psi_t(x)$. The resulting transport map is:

$$T_t(x) = \nabla \Phi_t(x) = x + \nabla \psi_t(x). \tag{29}$$

Computing the gradient of ψ_t , the explicit form of the map is:

$$T_t(x) = x + \varepsilon_t \sum_{j=1}^J w_{t,j} h_\beta(\langle a_{t,j}, x \rangle - c_{t,j}) a_{t,j},$$

where $h_\beta(z) := \zeta_\beta(z) \sigma_\beta(z)$. The parameter $\varepsilon_t > 0$ controls the magnitude of the deformation. For the pre-change process, we set $\varepsilon_t = 0.3$.

G.3. Verification: Cyclic Monotonicity and OT Optimality

The construction in equation 29 ensures that T_t is the gradient of a convex function. This property implies cyclic monotonicity, ensuring that if the reference law is absolutely continuous, T_t coincides with the unique quadratic-cost OT map from $\bar{\mu}$ to its pushforward μ_t .

Lemma G.1 (Convexity of the Potential). *For any t , the perturbation potential ψ_t is convex on \mathcal{X} . Consequently, the total potential $\Phi_t(x) = \frac{1}{2}\|x\|^2 + \psi_t(x)$ is strictly convex.*

Proof. The function $z \mapsto \zeta_\beta(z)$ is convex and non-decreasing. The function $y \mapsto y^2$ is convex and non-decreasing for $y \geq 0$. Since $\zeta_\beta(z) > 0$, the composition $z \mapsto (\zeta_\beta(z))^2$ is convex. The term $x \mapsto \langle a_{t,j}, x \rangle - c_{t,j}$ is affine. Since the composition of a convex function with an affine map preserves convexity, and a non-negative weighted sum of convex functions is convex, ψ_t is convex. Adding the strictly convex term $\frac{1}{2}\|x\|^2$ ensures Φ_t is strictly convex. \square

Lemma G.2 (Cyclic Monotonicity). *Let $\Phi : \mathcal{X} \rightarrow \mathbb{R}$ be convex and differentiable, and let $T = \nabla\Phi$. Then T is cyclically monotone: for any finite sequence $x_1, \dots, x_m \in \mathcal{X}$ with $x_{m+1} = x_1$,*

$$\sum_{i=1}^m \langle x_i, T(x_i) \rangle \geq \sum_{i=1}^m \langle x_i, T(x_{i+1}) \rangle.$$

Proof. By the first-order convexity condition, for any x, y , we have $\Phi(y) \geq \Phi(x) + \langle \nabla\Phi(x), y - x \rangle$. Rearranging terms yields $\langle \nabla\Phi(x), x \rangle - \langle \nabla\Phi(x), y \rangle \geq \Phi(x) - \Phi(y)$. Applying this inequality to pairs (x_i, x_{i+1}) and summing over $i = 1, \dots, m$, the right-hand side telescopes to 0, yielding the result. \square

Proposition G.3 (Optimality of the Constructed Map). *Assume $\bar{\mu}$ is absolutely continuous with respect to the Lebesgue measure on \mathcal{X} . Consider the quadratic cost $c(x, y) = \|x - y\|^2$. Let $\mu_t := (T_t)_\# \bar{\mu}$, where $T_t = \nabla\Phi_t$ with Φ_t convex. Then, T_t is the $\bar{\mu}$ -almost everywhere unique optimal transport map from $\bar{\mu}$ to μ_t .*

Proof. By Lemma G.1, T_t is the gradient of a convex function. By Brenier’s Theorem, the unique quadratic-cost optimal transport map from an absolutely continuous measure $\bar{\mu}$ to any target measure μ_t is characterized as the gradient of a convex function. Since T_t satisfies this characterization and pushes $\bar{\mu}$ to μ_t by construction, it is the optimal map. \square

G.4. Post-Change Scenarios

We simulate a single change-point at time κ . For $t \leq \kappa$ (pre-change), batches are generated using the reference base sample $X_{\text{base}}^{(0)}$ and random deformations T_t as described above. For $t > \kappa$ (post-change), we modify the *base generator* to induce a distributional shift, while maintaining the same deformation mechanism.

(S1) Barycenter Change. We generate a perturbed reference distribution $\bar{\mu}^{(1)}$ by perturbing the component parameters of the original family with strength δ_{loc} . We draw a new base sample $X_{\text{base}}^{(1)} \stackrel{\text{i.i.d.}}{\sim} \bar{\mu}^{(1)}$ once. For all $t > \kappa$, observations are generated as $X_t = T_t(X_{\text{base}}^{(1)})$. This scenario simulates a shift in the central tendency (mean/mode) of the stream.

(S2) Multimodal Reweighting. Let the reference $\bar{\mu}$ be a mixture with weights $\pi \in \Delta^{K-1}$. We perturb these weights to obtain $\pi^{(1)}$, where $\pi_k^{(1)} \propto \pi_k \eta_k$, with η_k drawn from a distribution controlled by perturbation strength δ_{mm} . We keep the component parameters fixed and sample $X_{\text{base}}^{(1)} \sim \bar{\mu}^{(1)}$ using the new weights. Post-change batches are generated as $X_t = T_t(X_{\text{base}}^{(1)})$. This scenario isolates changes in mass allocation across modes, which are often difficult to detect using low-order summary statistics.

(S3) **Copula Shift.** We modify the cross-dimensional dependence structure while approximately preserving the marginal distributions. Starting from the initial base sample $X_{\text{base}}^{(0)}$, we apply an Iman-Conover rank-preserving reordering to induce a target correlation ρ , resulting in a modified base sample $X_{\text{base}}^{(1)}$. Post-change batches are generated as $X_t = T_t(X_{\text{base}}^{(1)})$. This scenario targets dependence shifts that may be invisible to methods that monitor marginals independently.

H. Details on Synthetic Discrete Streams

This appendix details the generation processes for discrete distribution-valued streams and the specific formulations of the baseline detectors used for comparison.

H.1. Data Generation Processes

Stream of Poisson Counts. We simulate monitoring scenarios involving count data, where at each time step t , we observe a batch of N count-valued samples denoted as $X_t = \{x_{t,i}\}_{i=1}^N$ with $x_{t,i} \in \mathbb{Z}_{\geq 0}$. In the pre-change regime, samples are drawn independently from a standard Poisson distribution with rate λ_0 . To simulate the emergence of rare, extreme events in the post-change phase, we model the distribution as a mixture of the background process and a Dirac mass centered at a high value k^* . Specifically, for a mixing proportion $\alpha \in (0, 1)$, the post-change samples follow the mixture law $(1 - \alpha) \text{Pois}(\lambda_0) + \alpha \delta_{k^*}$. We also considered a heavy-tail mixture scenario, discussed in the additional results, where samples are drawn from a mixture of two Poisson sources with distinct rates, optionally calibrated such that the global mean remains matched to the pre-change baseline.

Ordered Categorical Drift. For data defined on a finite ordinal support $\mathcal{X} = \{1, \dots, M\}$ (with $M = 6$), we model gradual drifts that respect the underlying ordinal structure. Let the pre-change probability mass function be denoted by $p_0 \in \Delta^{M-1}$. We define a target "right-shifted" distribution p_{shift} wherein probability mass moves to adjacent higher categories. Formally, the shifted probability for category j is zero for the first category, equals $p_{0,j-1}$ for intermediate categories, and accumulates the mass of the top two original categories at the upper boundary $j = M$. The stream evolves via a linear interpolation $p_t = (1 - \gamma_t)p_0 + \gamma_t p_{\text{shift}}$, where the drift parameter γ_t ramps linearly from 0 to 1 following the change point.

H.2. Discrete Baseline Detectors

To establish comparative performance benchmarks for discrete data, we implemented two standard sequential change-point detectors adapted for count and categorical streams, respectively.

The first baseline is the **Poisson Cumulative Sum Detector** (adapted from the c -Chart), designed for monitoring shifts in total counts. This method relies on the property that the sum of independent Poisson variables is itself Poisson distributed. For a batch at time t , we calculate the sufficient statistic $S_t = \sum_{i=1}^N x_{t,i}$. Using the pre-change samples, we estimate the expected mean count $\bar{c} = \mathbb{E}[S_t]$. A detection threshold is established at three standard deviations from the mean, defining an acceptance region $[\max(0, \bar{c} - 3\sqrt{\bar{c}}), \bar{c} + 3\sqrt{\bar{c}}]$. An alarm is raised at time t if the aggregate count S_t falls outside this interval, signaling a significant deviation in the rate parameter.

The second baseline is the **Multinomial Max-Deviation Detector**, used for monitoring categorical data. This approach tracks the maximum deviation of any single class proportion from its expected baseline, effectively monitoring the L_∞ norm of the proportion shift. Let $\hat{p}_{t,j}$ be the empirical proportion of class j in the batch at time t , and let $p_{0,j}$ be the expected proportion estimated from reference data. We compute the standardized deviation score $z_{t,j}$ for each class by normalizing the difference $\hat{p}_{t,j} - p_{0,j}$ by the standard error $\sqrt{p_{0,j}(1 - p_{0,j})/N}$. The monitoring statistic is defined as the maximum absolute deviation across all categories, $Z_t = \max_j |z_{t,j}|$. A change is declared if Z_t exceeds a threshold h , which is empirically calibrated to satisfy the target in-control average run length (ARL_0).

I. Computational Acceleration Details

The geometric framework proposed in this work relies on the computation of Wasserstein barycenters and optimal transport maps, which can be computationally intensive in high-dimensional or large-sample regimes. To ensure the scalability of our online change-point detector, we adopt the following acceleration strategies.

1045 I.1. Efficient Barycenter Estimation via Normalizing Flows

1046 Estimating the reference Wasserstein barycenter $\bar{\mu}$ typically requires solving a large-scale optimization problem over the
 1047 space of probability measures. To accelerate this in high-dimensional settings ($d \geq 10$) and with large sample sizes,
 1048 we leverage the *Conditional Normalizing Flow* (CNF) framework proposed by Visentin & Cheridito (2026). Instead of
 1049 discretizing the support, this method parameterizes the barycenter and the associated transport maps using invertible neural
 1050 networks (normalizing flows).
 1051

1052 I.2. Entropic Regularization for OT

1054 For the computation of pairwise transport costs and maps in the monitoring phase, exact linear programming solvers (with cu-
 1055 bic complexity $O(N^3)$) become a bottleneck for streaming applications. We therefore adopt the Sinkhorn algorithm (Cuturi,
 1056 2013), which solves the entropically regularized optimal transport problem:
 1057

$$1058 \mathcal{L}_\varepsilon(\mu, \nu) = \inf_{\gamma \in \Pi(\mu, \nu)} \int c(x, y) d\gamma(x, y) + \varepsilon H(\gamma),$$

1061 where $H(\gamma)$ is the entropic regularization term. This formulation allows the optimal coupling to be computed via efficient
 1062 iterative matrix scaling (Sinkhorn-Knopp algorithm), reducing the complexity to approximately $O(N^2)$.
 1063

1064 J. Additional Experimental Results

1066 In this section, we provide the complete set of experimental results for the synthetic continuous streams. We evaluate
 1067 the performance across a wide range of configurations by varying the batch size $N \in \{50, 100, 300\}$ and the dimension
 1068 $d \in \{1, 5, 10, 50\}$. The results presented below include the full trade-off curves comparing ARL_1 against the ARL_0 for the
 1069 Barycenter Change, Multimodal Reweight and Copula Shift scenarios. Additionally, we report the Detection Rate versus
 1070 False Alarm Rate to assess the sensitivity of the detectors at varying thresholds. Finally, we provide detailed tabular results
 1071 for all the scenarios, reporting the mean detection delay with standard errors over 10 replications at a fixed calibration target.
 1072
 1073
 1074
 1075
 1076
 1077
 1078
 1079
 1080
 1081
 1082
 1083
 1084
 1085
 1086
 1087
 1088
 1089
 1090
 1091
 1092
 1093
 1094
 1095
 1096
 1097
 1098
 1099

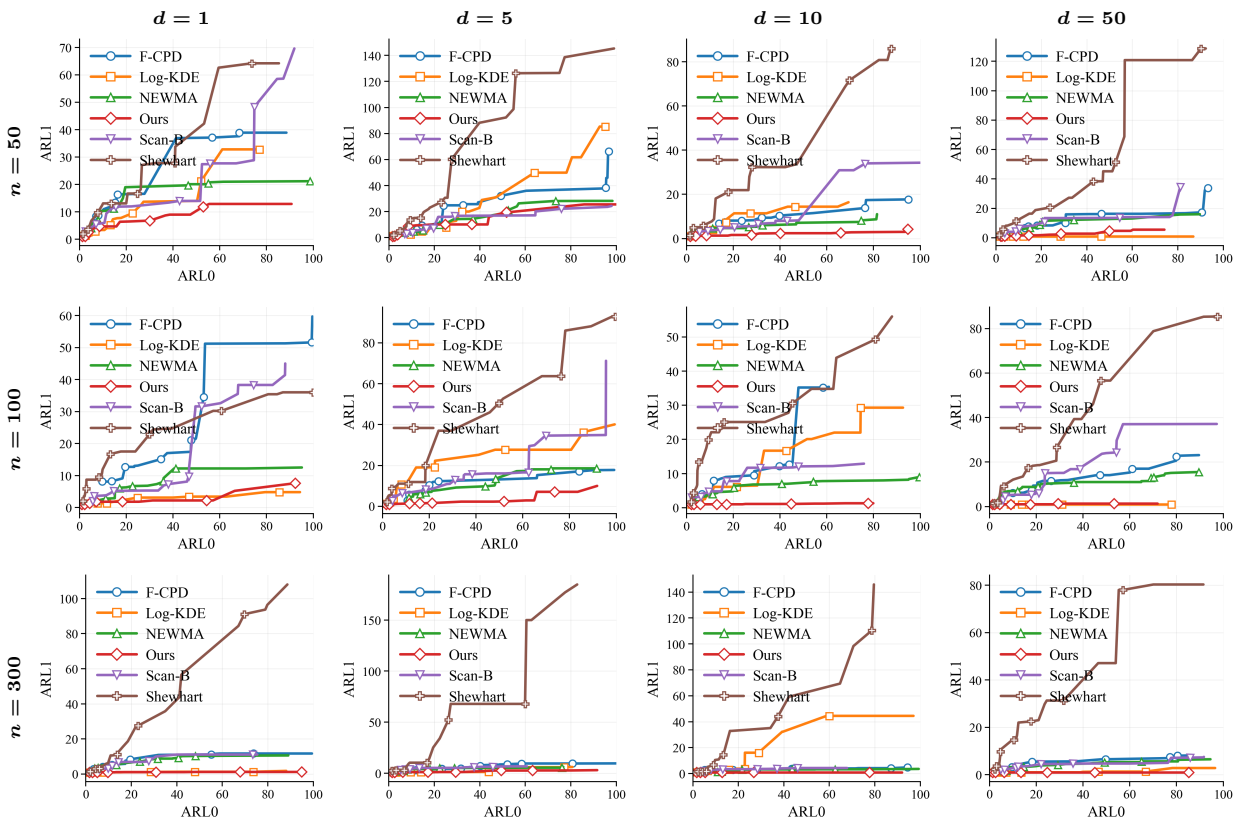


Figure 5. Multimodal Reweight (ARL_1 vs ARL_0): Performance comparison across varying sample sizes $n \in \{50, 100, 300\}$ (rows) and dimensions $d \in \{1, 5, 10, 50\}$ (columns).

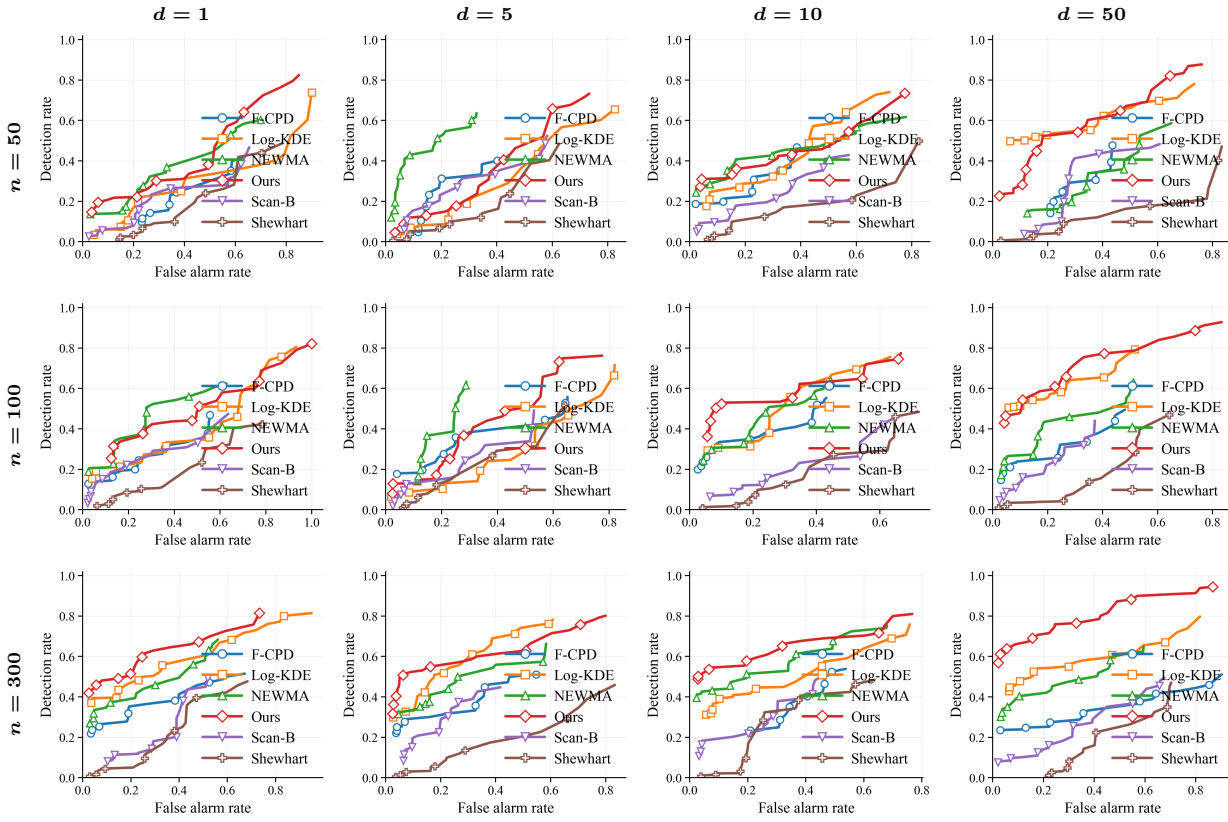


Figure 6. Multimodal Reweight (Detection Rate vs FAR): Performance comparison across varying sample sizes $n \in \{50, 100, 300\}$ (rows) and dimensions $d \in \{1, 5, 10, 50\}$ (columns).

Online CPD for Distribution-valued Data

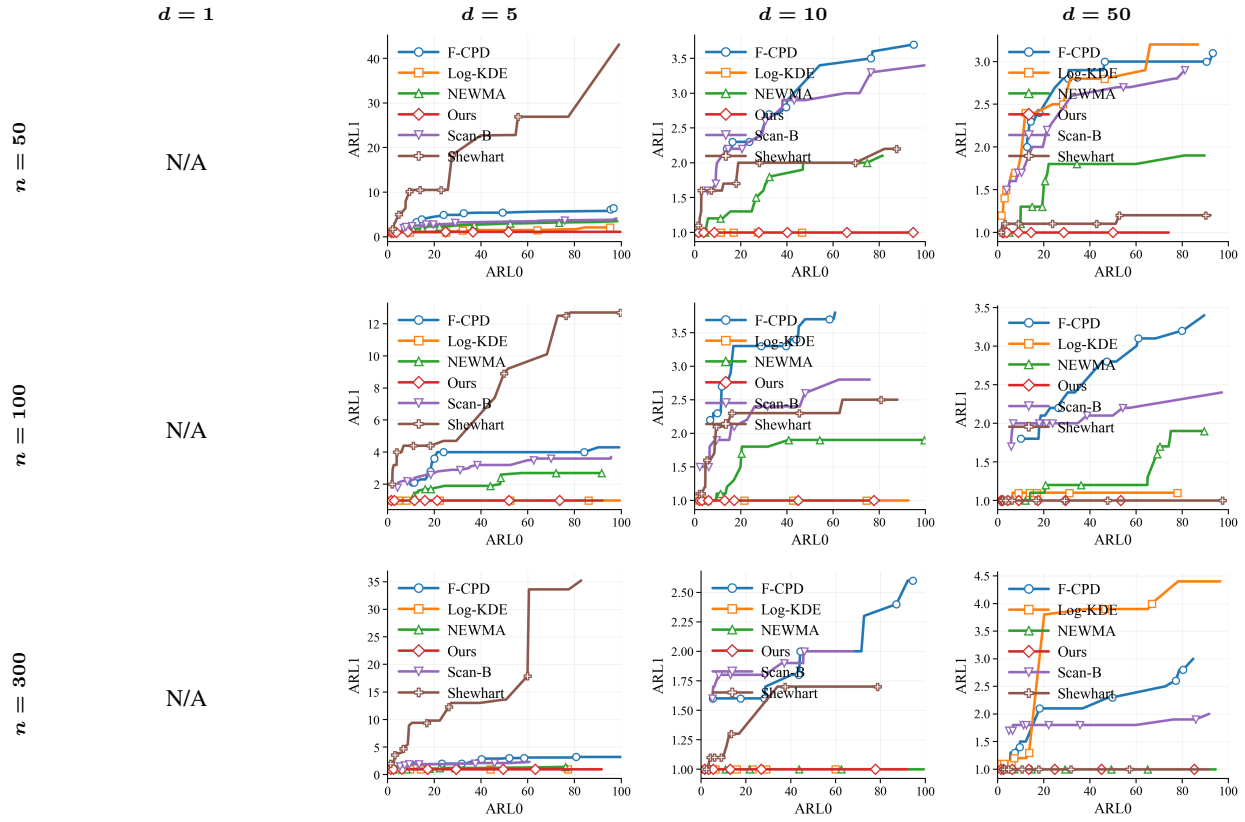


Figure 7. Copula Shift (ARL_1 vs ARL_0) Performance comparison across varying sample sizes $n \in \{50, 100, 300\}$ (rows) and dimensions $d \in \{5, 10, 50\}$ (columns).

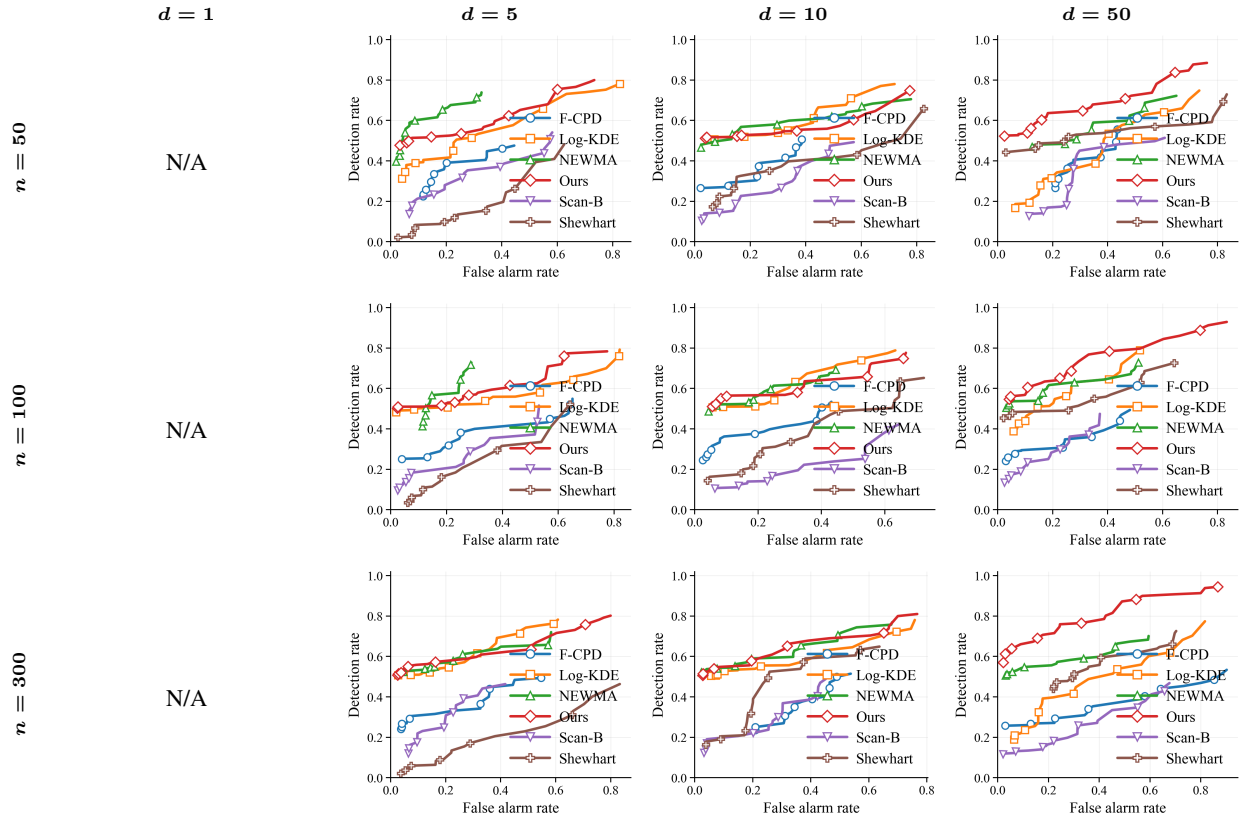


Figure 8. Copula Shift (Detection Rate vs FAR): Performance comparison across varying sample sizes $n \in \{50, 100, 300\}$ (rows) and dimensions $d \in \{5, 10, 50\}$ (columns).

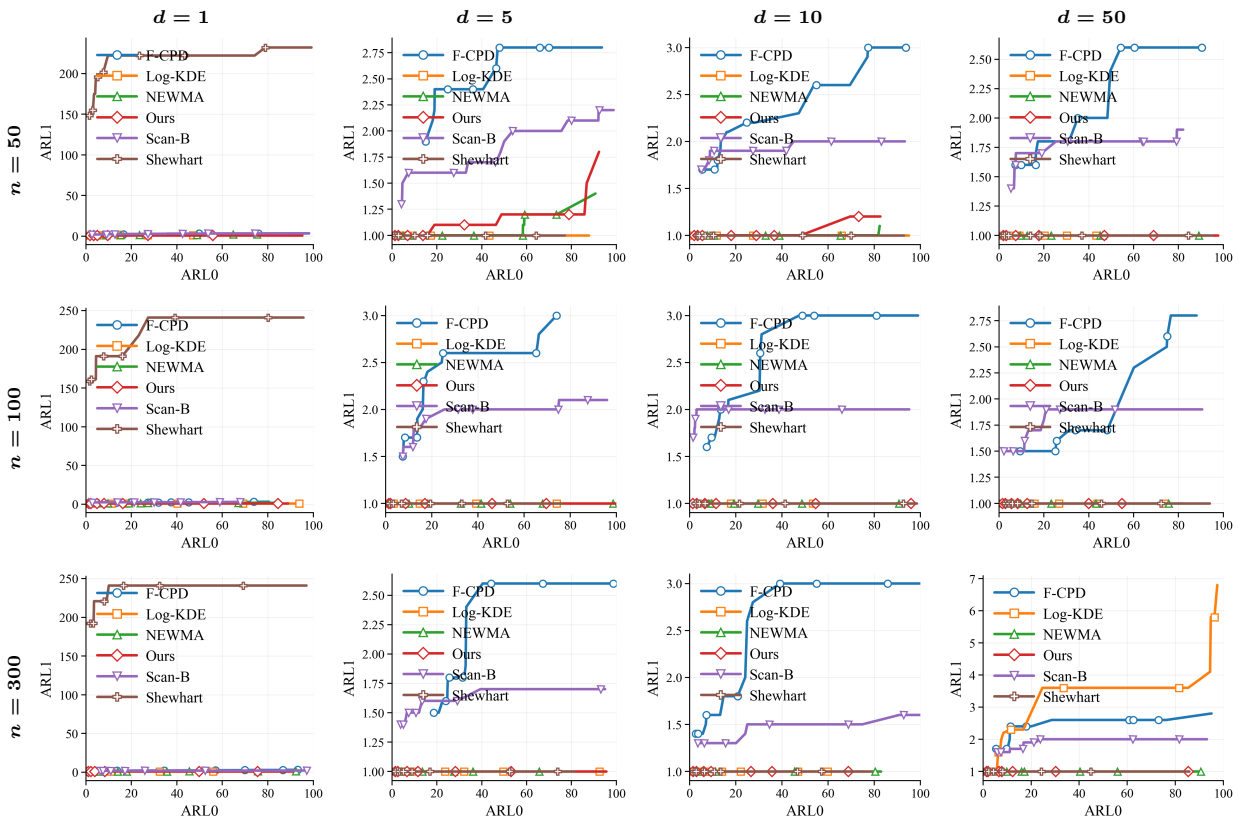


Figure 9. Barycenter Change (ARL_1 ARL_0): Performance comparison across varying sample sizes $n \in \{50, 100, 300\}$ (rows) and dimensions $d \in \{1, 5, 10, 50\}$ (columns).

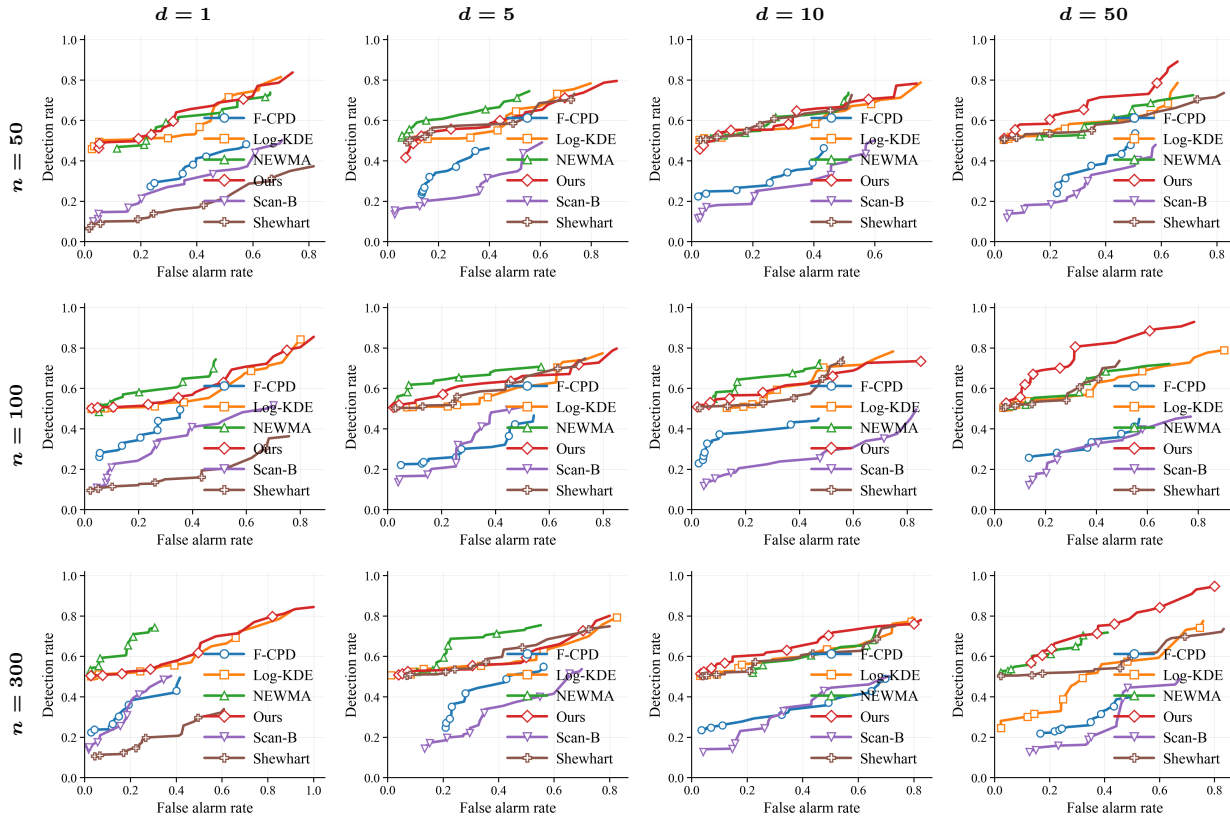


Figure 10. **Barycenter Change (Detection Rate vs FAR):** Performance comparison across varying sample sizes $n \in \{50, 100, 300\}$ (rows) and dimensions $d \in \{1, 5, 10, 50\}$ (columns).

J.1. Summary of Findings

The comprehensive results reinforce the robustness of the proposed framework across diverse shift types and dimensions. In the Barycenter scenario, where the shift is primarily in the mean, IDD achieves near-instantaneous detection ($ARL_1 \approx 1.0$), matching the performance of the specialized Shewhart chart. However, the advantage of our geometric approach becomes evident in complex distributional shifts such as Multimodal Reweighting. The Shewhart chart fails as the global mean remains unchanged, while IDD consistently maintains the lowest detection delay. Furthermore, regarding scalability, while density-based baselines like Log-KDE are competitive in lower dimensions, their performance degrades in high-dimensional settings ($d = 50$), particularly in the Copula Shift scenario. In contrast, IDD remains unaffected by the curse of dimensionality, outperforming most baselines in these challenging regimes.

K. Details on Flow Cytometry Case Study

K.1. Dataset and Preprocessing

The dataset is derived from the FlowCAP-II challenge (Aghaeepour et al., 2013), comprising 2,874 peripheral blood or bone marrow measurements collected from 359 subjects (316 healthy donors and 43 AML patients). Note that multiple measurements are available for each subject. We treat each flow cytometry measurement as a distinct distribution-valued observation. From each measurement, we extract a random subsample of $N = 2,000$ cells to form the empirical measure μ_t . Each cell is represented in $d = 7$ dimensions, consisting of forward scatter (FSC), side scatter (SSC), CD45 expression, and 4 other lineage markers.

Table 3. Synthetic results for **Barycenter** with $d = 1$ (10 replications). Best (smallest) ARL_1 is bold.

Method	N	ARL_0 (emp)	ARL_1
F-CPD	50	79.8	3.3 ± 0.7
Log-KDE	50	99.4	1.1 ± 0.1
NEWMA	50	101.6	2.4 ± 0.4
Ours	50	102.9	1.0 ± 0.0 ($\uparrow 9.1\%$)
Scan-B	50	101.1	3.4 ± 0.4
Shewhart	50	99.2	232.1 ± 39.5
<hr/>			
F-CPD	100	81.5	3.2 ± 0.3
Log-KDE	100	104.4	1.0 ± 0.0 ($\uparrow 47.4\%$)
NEWMA	100	111.9	1.9 ± 0.4
Ours	100	103.4	1.0 ± 0.0 ($\uparrow 47.4\%$)
Scan-B	100	101.2	2.9 ± 0.4
Shewhart	100	103.7	241.0 ± 40.0
<hr/>			
F-CPD	300	94.2	3.0 ± 0.0
Log-KDE	300	101.6	1.0 ± 0.0 ($\uparrow 23.1\%$)
NEWMA	300	105.5	1.3 ± 0.2
Ours	300	107.2	1.0 ± 0.0 ($\uparrow 23.1\%$)
Scan-B	300	97.1	2.1 ± 0.2
Shewhart	300	97.0	241.0 ± 40.0

Table 4. Synthetic results for **Barycenter** with $d = 5$ (10 replications). Best (smallest) ARL_1 is bold.

Method	N	ARL_0 (emp)	ARL_1
F-CPD	50	93.6	2.8 ± 0.2
Log-KDE	50	107.2	1.0 ± 0.0 ($\uparrow 28.6\%$)
NEWMA	50	90.8	1.4 ± 0.2
Ours	50	101.6	1.8 ± 0.5
Scan-B	50	98.7	2.2 ± 0.1
Shewhart	50	105.3	1.0 ± 0.0 ($\uparrow 28.6\%$)
<hr/>			
F-CPD	100	102.1	3.0 ± 0.0
Log-KDE	100	97.8	1.0 ± 0.0 ($\uparrow 52.4\%$)
NEWMA	100	98.8	1.0 ± 0.0 ($\uparrow 52.4\%$)
Ours	100	99.7	1.0 ± 0.0 ($\uparrow 52.4\%$)
Scan-B	100	96.0	2.1 ± 0.1
Shewhart	100	104.4	1.0 ± 0.0 ($\uparrow 52.4\%$)
<hr/>			
F-CPD	300	100.4	2.6 ± 0.3
Log-KDE	300	100.6	1.0 ± 0.0 ($\uparrow 41.2\%$)
NEWMA	300	100.7	1.0 ± 0.0 ($\uparrow 41.2\%$)
Ours	300	95.5	1.0 ± 0.0 ($\uparrow 41.2\%$)
Scan-B	300	95.0	1.7 ± 0.2
Shewhart	300	101.6	1.0 ± 0.0 ($\uparrow 41.2\%$)

Table 5. Synthetic results for **Barycenter** with $d = 10$ (10 replications). Best (smallest) ARL_1 is bold.

Method	N	ARL_0 (emp)	ARL_1
F-CPD	50	94.3	3.0 ± 0.0
Log-KDE	50	104.3	1.0 ± 0.0 ($\uparrow 16.7\%$)
NEWMA	50	100.8	1.2 ± 0.1
Ours	50	101.2	3.0 ± 1.9
Scan-B	50	93.3	2.0 ± 0.0
Shewhart	50	92.9	1.0 ± 0.0 ($\uparrow 16.7\%$)
<hr/>			
F-CPD	100	99.0	3.0 ± 0.0
Log-KDE	100	108.2	1.0 ± 0.0 ($\uparrow 50.0\%$)
NEWMA	100	101.1	1.0 ± 0.0 ($\uparrow 50.0\%$)
Ours	100	95.9	1.0 ± 0.0 ($\uparrow 50.0\%$)
Scan-B	100	100.5	2.0 ± 0.0
Shewhart	100	98.5	1.0 ± 0.0 ($\uparrow 50.0\%$)
<hr/>			
F-CPD	300	100.1	3.0 ± 0.0
Log-KDE	300	105.3	1.0 ± 0.0 ($\uparrow 37.5\%$)
NEWMA	300	100.2	1.0 ± 0.0 ($\uparrow 37.5\%$)
Ours	300	100.9	1.0 ± 0.0 ($\uparrow 37.5\%$)
Scan-B	300	99.8	1.6 ± 0.2
Shewhart	300	101.0	1.0 ± 0.0 ($\uparrow 37.5\%$)

Table 6. Synthetic results for **Barycenter** with $d = 50$ (10 replications). Best (smallest) ARL_1 is bold.

Method	N	ARL_0 (emp)	ARL_1
F-CPD	50	101.4	2.6 ± 0.3
Log-KDE	50	102.6	1.0 ± 0.0 ($\uparrow 50.0\%$)
NEWMA	50	104.0	1.0 ± 0.0 ($\uparrow 50.0\%$)
Ours	50	97.7	1.0 ± 0.0 ($\uparrow 50.0\%$)
Scan-B	50	101.3	2.0 ± 0.0
Shewhart	50	95.1	1.0 ± 0.0 ($\uparrow 50.0\%$)
<hr/>			
F-CPD	100	101.2	2.8 ± 0.2
Log-KDE	100	94.0	1.0 ± 0.0 ($\uparrow 47.4\%$)
NEWMA	100	95.1	1.0 ± 0.0 ($\uparrow 47.4\%$)
Ours	100	96.8	1.0 ± 0.0 ($\uparrow 47.4\%$)
Scan-B	100	107.2	1.9 ± 0.1
Shewhart	100	93.9	1.0 ± 0.0 ($\uparrow 47.4\%$)
<hr/>			
F-CPD	300	95.1	3.3 ± 0.3
Log-KDE	300	97.5	6.8 ± 2.9
NEWMA	300	90.6	1.0 ± 0.0 ($\uparrow 50.0\%$)
Ours	300	110.0	1.0 ± 0.0 ($\uparrow 50.0\%$)
Scan-B	300	104.0	2.0 ± 0.0
Shewhart	300	107.0	1.0 ± 0.0 ($\uparrow 50.0\%$)

Table 7. Synthetic results for **Copula Shift** with $d = 5$ (10 replications). Best (smallest) ARL_1 is bold.

Method	N	ARL_0 (emp)	ARL_1
F-CPD	50	96.8	6.6 ± 0.7
Log-KDE	50	103.2	2.3 ± 0.6
NEWMA	50	89.3	3.4 ± 0.2
Ours	50	100.1	1.1 ± 0.1 ($\uparrow 52.2\%$)
Scan-B	50	98.3	4.3 ± 0.2
Shewhart	50	99.0	43.1 ± 17.9
<hr/>			
F-CPD	100	102.0	4.4 ± 0.2
Log-KDE	100	99.3	1.0 ± 0.0 ($\uparrow 64.3\%$)
NEWMA	100	91.6	2.8 ± 0.2
Ours	100	91.8	1.0 ± 0.0 ($\uparrow 64.3\%$)
Scan-B	100	95.6	3.9 ± 0.2
Shewhart	100	99.6	12.7 ± 4.2
<hr/>			
F-CPD	300	99.9	3.2 ± 0.1
Log-KDE	300	116.7	1.0 ± 0.0 ($\uparrow 37.5\%$)
NEWMA	300	87.6	1.6 ± 0.2
Ours	300	91.6	1.0 ± 0.0 ($\uparrow 37.5\%$)
Scan-B	300	106.6	2.3 ± 0.2
Shewhart	300	96.3	35.6 ± 10.2

Table 8. Synthetic results for **Copula Shift** with $d = 10$ (10 replications). Best (smallest) ARL_1 is bold.

Method	N	ARL_0 (emp)	ARL_1
F-CPD	50	102.6	3.7 ± 0.3
Log-KDE	50	96.8	1.0 ± 0.0 ($\uparrow 52.4\%$)
NEWMA	50	112.5	2.1 ± 0.1
Ours	50	100.7	1.0 ± 0.0 ($\uparrow 52.4\%$)
Scan-B	50	99.9	3.4 ± 0.2
Shewhart	50	105.7	2.2 ± 0.5
<hr/>			
F-CPD	100	108.1	3.8 ± 0.1
Log-KDE	100	102.1	1.0 ± 0.0 ($\uparrow 47.4\%$)
NEWMA	100	99.6	1.9 ± 0.1
Ours	100	100.2	1.0 ± 0.0 ($\uparrow 47.4\%$)
Scan-B	100	113.1	2.8 ± 0.1
Shewhart	100	87.9	2.5 ± 0.4
<hr/>			
F-CPD	300	94.1	2.6 ± 0.3
Log-KDE	300	102.2	1.0 ± 0.0 ($\uparrow 44.4\%$)
NEWMA	300	100.7	1.0 ± 0.0 ($\uparrow 44.4\%$)
Ours	300	91.7	1.0 ± 0.0 ($\uparrow 44.4\%$)
Scan-B	300	100.1	2.0 ± 0.0
Shewhart	300	107.7	1.8 ± 0.3

Table 9. Synthetic results for **Copula Shift** with $d = 50$ (10 replications). Best (smallest) ARL_1 is bold.

Method	N	ARL_0 (emp)	ARL_1
F-CPD	50	101.4	2.6 ± 0.3
Log-KDE	50	102.6	1.0 ± 0.0 ($\uparrow 50.0\%$)
NEWMA	50	104.0	1.0 ± 0.0 ($\uparrow 50.0\%$)
Ours	50	97.7	1.0 ± 0.0 ($\uparrow 50.0\%$)
Scan-B	50	101.3	2.0 ± 0.0
Shewhart	50	95.1	1.0 ± 0.0 ($\uparrow 50.0\%$)
<hr/>			
F-CPD	100	101.2	2.8 ± 0.2
Log-KDE	100	94.0	1.0 ± 0.0 ($\uparrow 47.4\%$)
NEWMA	100	95.1	1.0 ± 0.0 ($\uparrow 47.4\%$)
Ours	100	96.8	1.0 ± 0.0 ($\uparrow 47.4\%$)
Scan-B	100	107.2	1.9 ± 0.1
Shewhart	100	93.9	1.0 ± 0.0 ($\uparrow 47.4\%$)
<hr/>			
F-CPD	300	95.1	3.3 ± 0.3
Log-KDE	300	97.5	6.8 ± 2.9
NEWMA	300	90.6	1.0 ± 0.0 ($\uparrow 50.0\%$)
Ours	300	110.0	1.0 ± 0.0 ($\uparrow 50.0\%$)
Scan-B	300	104.0	2.0 ± 0.0
Shewhart	300	107.0	1.0 ± 0.0 ($\uparrow 50.0\%$)

Table 10. Synthetic results for **Multimodal Reweight** with $d = 1$ (10 replications). Best (smallest) ARL_1 is bold.

Method	N	ARL_0 (emp)	ARL_1
F-CPD	50	103.0	49.6 ± 21.6
Log-KDE	50	108.3	45.6 ± 15.6
NEWMA	50	107.9	24.1 ± 5.4
Ours	50	107.7	15.7 ± 9.4 ($\uparrow 34.9\%$)
Scan-B	50	93.2	69.7 ± 21.0
Shewhart	50	101.1	93.5 ± 28.1
<hr/>			
F-CPD	100	99.5	59.7 ± 24.6
Log-KDE	100	94.3	4.8 ± 1.7 ($\uparrow 36.8\%$)
NEWMA	100	90.1	12.2 ± 4.1
Ours	100	92.3	7.6 ± 4.0
Scan-B	100	106.2	70.0 ± 31.2
Shewhart	100	99.9	36.0 ± 8.7
<hr/>			
F-CPD	300	99.4	11.8 ± 2.4
Log-KDE	300	88.1	1.8 ± 0.5
NEWMA	300	105.8	10.6 ± 2.6
Ours	300	94.9	1.2 ± 0.2 ($\uparrow 33.3\%$)
Scan-B	300	109.5	11.1 ± 4.2
Shewhart	300	94.5	108.0 ± 20.6

Table 11. Synthetic results for **Multimodal Reweight** with $d = 5$ (10 replications). Best (smallest) ARL_1 is bold.

Method	N	ARL_0 (emp)	ARL_1
F-CPD	50	96.8	66.2 ± 29.4
Log-KDE	50	103.2	85.4 ± 34.4
NEWMA	50	89.3	28.2 ± 6.1
Ours	50	100.1	25.5 ± 9.5 ($\uparrow 46.9\%$)
Scan-B	50	98.3	27.4 ± 8.6
Shewhart	50	99.0	145.3 ± 36.6
F-CPD	100	102.0	21.1 ± 4.3
Log-KDE	100	99.3	40.1 ± 14.7
NEWMA	100	91.6	18.6 ± 6.8
Ours	100	91.8	10.0 ± 4.4 ($\uparrow 46.2\%$)
Scan-B	100	95.6	72.3 ± 32.6
Shewhart	100	99.6	92.9 ± 26.6
F-CPD	300	99.9	9.7 ± 1.0
Log-KDE	300	116.7	6.3 ± 4.5
NEWMA	300	87.6	6.5 ± 0.8
Ours	300	91.6	3.1 ± 1.1 ($\uparrow 50.8\%$)
Scan-B	300	106.6	6.8 ± 0.6
Shewhart	300	96.3	190.8 ± 40.5

Table 12. Synthetic results for **Multimodal Reweight** with $d = 10$ (10 replications). Best (smallest) ARL_1 is bold.

Method	N	ARL_0 (emp)	ARL_1
F-CPD	50	102.6	17.7 ± 5.2
Log-KDE	50	96.8	16.4 ± 10.2
NEWMA	50	112.5	11.3 ± 3.4
Ours	50	100.7	4.3 ± 1.9 ($\uparrow 61.9\%$)
Scan-B	50	99.9	34.3 ± 17.9
Shewhart	50	105.7	87.3 ± 25.2
F-CPD	100	108.1	35.4 ± 22.2
Log-KDE	100	102.1	29.3 ± 18.2
NEWMA	100	99.6	10.6 ± 2.7
Ours	100	100.2	1.5 ± 0.4 ($\uparrow 85.8\%$)
Scan-B	100	113.1	13.0 ± 3.6
Shewhart	100	87.9	56.0 ± 9.2
F-CPD	300	94.1	4.8 ± 1.0
Log-KDE	300	102.2	44.5 ± 30.7
NEWMA	300	100.7	3.7 ± 0.8
Ours	300	91.7	1.0 ± 0.0 ($\uparrow 73.0\%$)
Scan-B	300	100.1	4.4 ± 0.7
Shewhart	300	107.7	145.8 ± 35.1

K.2. Change Point Detection Experiment Setup

We structure the experiment as a sequential change-point detection task with two phases: (1) Pre-change Calibration Phase: We utilize a sequence of 300 healthy samples to estimate the reference distribution $\bar{\mu}$ and calibrate the detection threshold to satisfy a target ARL_0 . (2) Monitoring Phase: We monitor a test stream of 300 samples. This stream is constructed by injecting 80% AML-positive samples (randomly sampled from the AML patient pool) interspersed with healthy samples. Since ground truth diagnosis labels are available for every sample, we evaluate performance using both detection delay (ARL_1) and classification metrics (F1-score, Precision/Recall) at the batch level.

L. Details on Reddit Vaccine Sentiment Study

L.1. Preprocessing and Embedding Pipeline

To transform the unstructured text stream into a sequence of distribution-valued observations suitable for our geometric CPD framework, we applied the following preprocessing steps: Firstly, we aggregated user comments into daily batches ($t = 1, \dots, T$). To ensure statistical stability of the empirical measures, we filtered out days containing fewer than 30 comments. Secondly, we mapped each raw comment to a dense semantic vector using the pre-trained Sentence-BERT model (all-MiniLM-L6-v2 (Reimers & Gurevych, 2019)), producing 384-dimensional embeddings. This places semantically similar comments (e.g., expressing anxiety or support) close to each other in vector space. Finally, to reduce computational cost while preserving the primary modes of variation, we projected the embeddings to $d = 20$ dimensions using Principal Component Analysis (PCA).

L.2. Timeline of Interest

Table 13 details the key external events overlapping with our monitoring window. The qualitative evaluation in the main text assesses whether detection alarms coincide with the onset of public reaction (polarization or anxiety) triggered by these events.

1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649

Table 13. Timeline of Key COVID-19 Vaccine Events (Monitoring Phase).

Date	Event	Description
<i>Feb 27</i>	J&J EUA	FDA issues Emergency Use Authorization for Johnson & Johnson vaccine. (Start of monitoring).
<i>Mar 08</i>	CDC Guidance	CDC releases initial guidance for fully vaccinated individuals.
<i>Apr 13</i>	J&J Pause	FDA/CDC recommend pausing J&J administration due to rare clotting reports. (Primary shock event).
<i>Apr 19</i>	Eligibility	All U.S. adults become eligible for vaccination.
<i>Apr 23</i>	Pause Lifted	CDC safety panel recommends resuming J&J vaccinations; pause lifted.
<i>Apr 27</i>	Mask Relaxation	CDC relaxes outdoor mask guidelines for vaccinated people.
<i>May 04</i>	July 4th Goal	President Biden announces 70% vaccination goal by July 4th.